

STATISTICAL RESEARCH REPORT
Institute of Mathematics
University of Oslo

No 7
October 1978

ON MODELS AND METHODS
OF CREDIBILITY

by

Bjørn Sundt

ACKNOWLEDGEMENTS.

I am grateful to Ragnar Norberg for his ideas and encouragement. I would also like to thank the Library of the Institute of Actuaries, London, for providing me with material on credibility theory.

CONTENTS.

	Page
0. Introduction.	1
1. Credibility models with applications to regression.	3
2. An insurance model with collective seasonal random factors.	28
3. A hierarchical credibility regression model.	40
4. Optimal choice of observators.	50
5. Optimal semilinear estimators.	62
References.	73

0. Introduction.

0A. Credibility theory originated in experience ratemaking in insurance. Suppose that an insurance policy was originally rated with a manual premium p . After some time we get some claim experience from this policy (claim numbers and claim sizes). This experience tells us something about the risk properties of this policy and accordingly ought to be used to adjust the premium. A common premium to use is

$$(0.1) \quad z \bar{x} + (1 - z) p.$$

Here, \bar{x} is some estimate, based on our experience data, of the expected total claim amount per insurance term. The constant z tells us how much weight to put on our experience, that is, how credible it is. z is therefore called the credibility coefficient or sometimes more briefly the credibility. A premium of the form (0.1) was first developed by Whitney (1918).

About 1970 one started to develop more complicated formulae, e.g. of the form

$$\gamma_0 + \gamma_1 \bar{x}_1 + \gamma_2 \bar{x}_2,$$

where \bar{x}_1 (\bar{x}_2) is the observed mean of the claim numbers (sizes) per insurance term (Hewitt (1970) with discussion).

In the present thesis (as in most modern credibility theory), we shall use the expression credibility estimators for estimators that are in some wide sense linear. We shall mainly be interested in estimating unknown random variables.

A survey of classical credibility theory is given by Longley-Cook (1962) and one of modern credibility theory by Jewell (1976). Both of these papers contain an extensive list of references.

0B. The present thesis consists of five distinct sections, each of which starts with an introduction, giving the main ideas of the section and connecting these to existing theory.

0C. Notation. Matrices and vectors are written in doubly underlined letters, capital and lower case respectively, e.g. A and a. No notational distinction is made between random and nonrandom quantities, except that Greek letters are reserved for parameters. A few random variables are denoted by Greek letters, as they could be interpreted as parameters in Bayesian sense.

Identity matrices are denoted by I and matrices containing only zeroes by 0. The dimensions of such matrices will not be given explicitly, as they will be clear from the expressions in which the matrices appear.

Let x = (x₁, ..., x_r)' and u = (u₁, ..., u_s)' be two random vectors.

E(x) denotes the r × 1-vector whose i-th element is the expectation of x_i.

C(x, u') denotes the r × s-matrix whose (i,j)-element is the covariance between x_i and u_j. The covariance matrix C(x, x') of x is denoted by C(x).

If θ is a random variable, E(x|θ), C(x, u'|θ), and C(x|θ) correspond to the above definitions in the conditional distribution given θ.

The symbol □ is used to indicate the end of a section of examples.

0D. All displayed expectations and covariances are assumed to exist.

1. Credibility models with applications to regression.

A general regression model is given from which models by Hachemeister (1975), Taylor (1977), and Jewell (1975a,b) drop out as special cases. The connection between homogeneous and inhomogeneous estimators is analyzed, and a new interpretation of best linear unbiased homogeneous estimators is given. Concepts of unbiasedness and θ -unbiasedness give interpretations of the credibility estimators. A concept of exchangeability can sometimes be used to simplify the calculation of the estimators. Finally we discuss what happens if we replace the constant term in an inhomogeneous estimator by an "old estimator".

1A. Let $\underline{x} = (x_1, \dots, x_r)'$ be an observable random vector and m an unknown random variable. We want to estimate m by an estimator \dot{m} from a certain set of estimators based on \underline{x} . When \dot{m} is a linear function of \underline{x} , we shall call \dot{m} a linear estimator (based on \underline{x}).

Let $\dot{m}^{(1)}$ and $\dot{m}^{(2)}$ be two estimators of m . We shall say that $\dot{m}^{(1)}$ is better than $\dot{m}^{(2)}$ if

$$E(\dot{m}^{(1)} - m)^2 < E(\dot{m}^{(2)} - m)^2$$

(that is, we use quadratic loss).

Suppose now that we want to estimate a random vector $\underline{m} = (m_1, \dots, m_s)'$ and have two estimators $\underline{\dot{m}}^{(1)} = (\dot{m}_1^{(1)}, \dots, \dot{m}_s^{(1)})'$ and $\underline{\dot{m}}^{(2)} = (\dot{m}_1^{(2)}, \dots, \dot{m}_s^{(2)})'$. We shall say that $\underline{\dot{m}}^{(1)}$ is better than $\underline{\dot{m}}^{(2)}$ if $\dot{m}_i^{(1)}$ is not a worse estimator of m_i than $\dot{m}_i^{(2)}$ for all i and better for at least one i .

1B. We want to estimate m by \tilde{m} , the best linear inhomogeneous estimator (based on \underline{x}), that is, the best estimator of the form $g_0 + \sum_{i=1}^r g_i x_i$, where g_0, g_1, \dots, g_r are real constants.

Putting the derivatives of $E(g_0 + \sum_{i=1}^r g_i x_i - m)^2$ with respect to g_0, g_1, \dots, g_r equal to zero gives that the optimal values $\gamma_0, \gamma_1, \dots, \gamma_r$ of g_0, g_1, \dots, g_r must satisfy

$$(1.1) \quad \gamma_0 + \sum_{i=1}^r \gamma_i E(x_i) = E(m)$$

$$(1.2) \quad \gamma_0 E(x_j) + \sum_{i=1}^r \gamma_i E(x_i x_j) = E(m x_j) \quad j=1, \dots, r,$$

By multiplying (1.1) by $E(x_j)$ and subtracting from (1.2) we get

$$(1.3) \quad \sum_{i=1}^r \gamma_i C(x_i, x_j) = C(m, x_j) \quad j=1, \dots, r.$$

If we let $\underline{\Sigma} = C(\underline{x})$ and $\underline{\gamma} = (\gamma_1, \dots, \gamma_r)'$, our linear system can be written on the matrix form

$$(1.4) \quad \gamma_0 + \underline{\gamma}' E(\underline{x}) = E(m)$$

$$(1.5) \quad \underline{\gamma}' \underline{\Sigma} = C(m, \underline{x}')$$

We assume $\underline{\Sigma}$ invertible. Then we have the unique solution

$$(1.6) \quad \underline{\gamma}' = C(m, \underline{x}') \underline{\Sigma}^{-1}$$

$$(1.7) \quad \gamma_0 = E(m) - C(m, \underline{x}') \underline{\Sigma}^{-1} E(\underline{x})$$

Consequently

$$(1.8) \quad \tilde{m} = C(m, \underline{x}') \underline{\Sigma}^{-1} [\underline{x} - E(\underline{x})] + E(m).$$

The best linear inhomogeneous estimator of the random vector \underline{m} is of course

$$\tilde{\underline{m}} = C(\underline{m}, \underline{x}') \underline{\Sigma}^{-1} [\underline{x} - E(\underline{x})] + E(\underline{m}).$$

1C. If r is large, constructing the estimators of the previous subsection can involve lots of work in inverting and multiplying. It is therefore of importance to find some simplifications.

The following concept of exchangeability is often useful in this respect.

Definition 1.1. Let $\underline{u}, \underline{x}_1, \underline{x}_2, \dots, \underline{x}_n$ be random vectors, $\underline{x}_1, \dots, \underline{x}_n$ of same dimension. We shall say that $\underline{x}_1, \dots, \underline{x}_n$ are exchangeable relative to \underline{u} if $(\underline{x}_{i_1}, \dots, \underline{x}_{i_n}, \underline{u})$ has the same joint distribution for all permutations (i_1, \dots, i_n) of $(1, \dots, n)$.

We note that if $\underline{x}_1, \dots, \underline{x}_n$ are exchangeable relative to \underline{u} , they are also exchangeable in De Finetti's sense, but that the converse implication is not always true. Hence the present concept of exchangeability is stronger than De Finetti's.

Let m be an unknown random variable and $\underline{x}_0, \underline{x}_1, \dots, \underline{x}_n$ observable random vectors, $\underline{x}_1, \dots, \underline{x}_n$ of same dimension. Let

$$\tilde{m} = \gamma + \sum_{i=0}^n \underline{\gamma}_i' \underline{x}_i$$

be the best linear inhomogeneous estimator of m based on $\underline{x}_0, \underline{x}_1, \dots, \underline{x}_n$. $\gamma, \underline{\gamma}_0, \dots, \underline{\gamma}_n$ are assumed to be uniquely determined.

Theorem 1.1. If $\underline{x}_1, \dots, \underline{x}_n$ are exchangeable relative to $(\underline{x}_0, m)'$, then $\underline{\gamma}_1 = \underline{\gamma}_2 = \dots = \underline{\gamma}_n$.

Proof. Let $k \in \{1, \dots, n-1\}$. Then

$$E(\gamma + \sum_{i=0}^n \gamma_i' x_i - m)^2 = E(\gamma + \sum_{\substack{i=0 \\ i \neq k}}^{n-1} \gamma_i' x_i + \gamma_k' x_k + \gamma_n' x_n - m)^2 =$$

$$E(\gamma + \sum_{\substack{i=0 \\ i \neq k}}^{n-1} \gamma_i' x_i + \gamma_k' x_n + \gamma_n' x_k - m)^2 =$$

$$E(\gamma + \sum_{\substack{i=0 \\ i \neq k}}^{n-1} \gamma_i' x_i + \gamma_n' x_k + \gamma_k' x_n - m)^2.$$

As $\gamma, \gamma_0, \gamma_1, \dots, \gamma_n$ were assumed unique, we must have $\gamma_k = \gamma_n$. And since k was arbitrary, $\gamma_1 = \gamma_2 = \dots = \gamma_n$.

Q E D.

From Theorem 1.1 now follows that \tilde{m} is the best estimator of m of the form

$$g + g_0' x_0 + g' \bar{x},$$

where $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$, where g is a non-random number and g_0 and g are non-random vectors.

1D. In the model of subsection 1B assume that

$E(x) = Y \beta$ and $E(m) = a' \beta$, where β and a are non-random $q \times 1$ -vectors and Y is a non-random $r \times q$ -matrix. We assume that $\text{rank}(Y) = q$.

Under these conditions (1.8) becomes

$$(1.9) \quad \tilde{m} = C(m, x') \Sigma^{-1} x + [a' - C(m, x') \Sigma^{-1} Y] \beta.$$

We shall say that an estimator \hat{m} of m is unbiased (for all $\underline{\beta}$) if $E(\hat{m}) = E(m)$ for all values in \mathbb{R}^q of the parameter vector $\underline{\beta}$, that is, $E(m) = \underline{a}' \underline{\beta}$ for all values of $\underline{\beta}$.

For all linear unbiased homogeneous estimators $\underline{c}' \underline{x}$ of m we must have $\underline{a}' \underline{\beta} = E(m) = \underline{c}' E(\underline{x}) = \underline{c}' \underline{Y} \underline{\beta}$ for all $\underline{\beta} \in \mathbb{R}^q$, which is equivalent with

$$(1.10) \quad \underline{c}' \underline{Y} = \underline{a}'.$$

We get

$$\begin{aligned} E(\underline{c}' \underline{x} - m)^2 &= C(\underline{c}' \underline{x} - m) = C((\underline{y}' \underline{x} - m) + (\underline{c} - \underline{y})' \underline{x}) = \\ &= C(\underline{y}' \underline{x} - m) + C((\underline{c} - \underline{y})' \underline{x}) + 2 C(\underline{y}' \underline{x} - m, \underline{x}' (\underline{c} - \underline{y})) = \\ &= C(\underline{y}' \underline{x} - m) + C((\underline{c} - \underline{y})' \underline{x}) + 2[\underline{y}' \underline{\Sigma} - C(m, \underline{x}')] (\underline{c} - \underline{y}). \end{aligned}$$

From (1.5) follows that the last term is equal to zero, giving

$$E(\underline{c}' \underline{x} - m)^2 = C(\underline{y}' \underline{x} - m) + C(\underline{d}' \underline{x}),$$

where $\underline{d} = \underline{c} - \underline{y}$. Hence the best linear unbiased homogeneous estimator of m based on \underline{x} must be

$$\hat{m} = \underline{y}' \underline{x} + \underline{\delta}' \underline{x},$$

where $\underline{\delta}$ is the vector minimizing $C(\underline{d}' \underline{x})$ under the side condition

$$\underline{d}' \underline{Y} = \underline{a}' - C(m, \underline{x}') \underline{\Sigma}^{-1} \underline{Y}.$$

But then $\underline{\delta}' \underline{x}$ must be the best linear unbiased estimator of $[\underline{a}' - C(m, \underline{x}') \underline{\Sigma}^{-1} \underline{Y}] \underline{\beta}$. From standard least squares theory follows that

$$\underline{\delta}' \underline{x} = [\underline{a}' - C(m, \underline{x}') \underline{\Sigma}^{-1} \underline{Y}] \hat{\underline{\beta}},$$

where

$$(1.11) \quad \hat{\underline{\beta}} = (\underline{Y}' \underline{\Sigma}^{-1} \underline{Y})^{-1} \underline{Y}' \underline{\Sigma}^{-1} \underline{x}.$$

This gives

$$(1.12) \quad \ddot{m} = C(m, \underline{x}') \underline{\Sigma}^{-1} \underline{x} + [\underline{a}' - C(m, \underline{x}') \underline{\Sigma}^{-1} \underline{y}] \hat{\underline{\beta}}.$$

If $E(\underline{m}) = \underline{A} \underline{\beta}$, where \underline{A} is an $s \times q$ -matrix, the vector generalizations of (1.9) and (1.12) follow easily,

$$(1.13) \quad \tilde{\underline{m}} = C(\underline{m}, \underline{x}') \underline{\Sigma}^{-1} \underline{x} + [\underline{A} - C(\underline{m}, \underline{x}') \underline{\Sigma}^{-1} \underline{y}] \underline{\beta}$$

$$(1.14) \quad \ddot{\underline{m}} = C(\underline{m}, \underline{x}') \underline{\Sigma}^{-1} \underline{x} + [\underline{A} - C(\underline{m}, \underline{x}') \underline{\Sigma}^{-1} \underline{y}] \hat{\underline{\beta}}.$$

(1.13) and (1.14) give

Theorem 1.2. The best linear unbiased homogeneous estimator of \underline{m} is the same as the best linear inhomogeneous estimator of \underline{m} with the parameter vector $\underline{\beta}$ replaced by its best linear unbiased estimator,

A similar result has been shown by De Vylder (1978a),

1E. The historical development of homogeneous formulae in credibility theory seems, in short, to have been the following:

i) Bühlmann & Straub (1970) had a model where

$E(m) = E(x_1) = \dots = E(x_r) = \beta > 0$. They first developed \tilde{m} ,

in which the expectation β appears. But β was assumed unknown and had to be estimated. They then sought the best estimator of

m of the form $\dot{m} = \sum_{i=1}^r c_i x_i$ such that $E(m) = \beta$, that is,

$\beta \sum_{i=1}^r c_i = \beta$. Thus they had to minimize $E(\sum_{i=1}^r c_i x_i - m)^2$ under the side condition

$$(1.15) \quad \sum_{i=1}^r c_i = 1.$$

Compared with \tilde{m} this estimator has a built-in estimator of β , and from Theorem 1.2 follows that this built-in estimator is

a natural one. Note that we get the side condition (1.15) both if we require our estimator to be unbiased for one specific value of β , and if we require unbiasedness for a greater set of β 's.

ii) Hachemeister (1975) generalized Bühlmann and Straub's result to a multi-dimensional $\underline{\beta}$ by minimizing $E(\underline{c}'\underline{x} - m)^2$ under the side condition

$$\underline{a}'\underline{\beta} = \underline{c}'\underline{Y} \underline{\beta}$$

for a fixed $\underline{\beta}$. As $\underline{\beta}$ appeared in the resulting estimator, he concluded that in the multi-dimensional case there is no benefit in using homogeneous estimators.

iii) The present unbiasedness condition giving the side condition (1.10) was developed by Taylor, first in a generalization of Bühlmann and Straub's model with $q = 2$ (Taylor (1975)), and later in the general case (Taylor (1977)).

To the present author the following seems to be the most natural reasoning leading to minimization of $E(\underline{c}'\underline{x} - m)^2$ under side condition (1.10):

We developed \tilde{m} as the best estimator of m of the form $g_0 + \underline{g}'\underline{x}$, where g_0 is a non-random number and \underline{g} a non-random $r \times 1$ -vector. In this estimator $\underline{\beta}$ appears. If $\underline{\beta}$ is unknown, it would be natural to develop the best estimator of the form $g_0 + \underline{g}'\underline{x}$ that is unbiased for all $\underline{\beta}$. If $g_0 + \underline{g}'\underline{x}$ is unbiased for all $\underline{\beta}$, we must have

$$\underline{a}'\underline{\beta} = g_0 + \underline{g}'\underline{Y} \underline{\beta}$$

for all values of $\underline{\beta}$ in \mathbb{R}^q . This gives

$$g_0 = 0$$

and

$$(1.16) \quad \underline{g}'\underline{Y} = \underline{a}'.$$

Hence the class of linear, unbiased for all $\underline{\beta}$, estimators of m is the same as the class of linear unbiased homogeneous estimators of m .

After these considerations we could call \tilde{m} the best linear estimator of m (based on \underline{x}) and \hat{m} the best linear, unbiased for all $\underline{\beta}$, estimator of m (based on \underline{x}). However, we shall stick to the terminology already introduced.

1F. We shall now assume that \underline{x} and m are independent given an unknown random variable (possibly vector) θ . For any estimator \hat{m} of m based on \underline{x} we have

$$(1.17) \quad E(\hat{m} - m)^2 = E C(m|\theta) + E(\hat{m} - E(m|\theta))^2,$$

so that \hat{m} is an optimal estimator of m if and only if \hat{m} is an optimal estimator of $E(m|\theta)$.

We shall make the further assumptions

$$E(m|\theta) = \underline{a}'\underline{b}(\theta)$$

$$E(\underline{b}(\theta)) = \underline{\beta}$$

We assume that

$$\underline{A} = C(\underline{b}(\theta))$$

is invertible.

Similarly we assume that \underline{m} and \underline{x} are independent given θ and that

$$E(\underline{m}|\theta) = \underline{A} \underline{b}(\theta).$$

Corresponding to (1.17) we have

$$(1.18) \quad E((\hat{m} - \underline{m})(\hat{m} - \underline{m})') = E C(\underline{m}|\theta) + E((\hat{m} - E(\underline{m}|\theta))(\hat{m} - E(\underline{m}|\theta))').$$

The assumptions made in this subsection are very common in credibility theory and will be used in most parts of the present paper.

1G. We assume that

$$E(\underline{x}|\theta) = \underline{Y} \underline{b}(\theta)$$

$$EC(\underline{x}|\theta) = \underline{\Phi}.$$

Then we have

$$\underline{\Sigma} = \underline{\Phi} + \underline{Y} \underline{\Lambda} \underline{Y}'$$

$$(1.19) \quad C(\underline{m}, \underline{x}') = \underline{A} \underline{\Lambda} \underline{Y}'.$$

From (1.11), (1.13), and (1.19) follows

$$(1.20) \quad \underline{\tilde{m}} = \underline{A} [\underline{Z} \hat{\underline{\beta}} + (\underline{I} - \underline{Z}) \underline{\beta}]$$

where

$$\underline{Z} = \underline{A} \underline{Y}' \underline{\Sigma}^{-1} \underline{Y}.$$

Formula (1.20) was first shown by Hachemeister (1975).

Later contributions include Taylor (1977), Jewell (1975a,b), and De Vylder (1976a).

Theorem 1.2 and (1.20) now give

$$(1.21) \quad \underline{\ddot{m}} = \underline{A} \hat{\underline{\beta}}.$$

This simple formula will be discussed in subsection 1K.

By letting $\underline{m} = \underline{b}(\theta)$ we get

$$(1.22) \quad \underline{\tilde{b}} = \underline{Z} \hat{\underline{\beta}} + (\underline{I} - \underline{Z}) \underline{\beta}$$

$$(1.23) \quad \underline{\ddot{b}} = \hat{\underline{\beta}}.$$

Putting (1.22) and (1.23) into (1.20) and (1.21) gives

$$(1.24) \quad \underline{\tilde{m}} = \underline{A} \underline{\tilde{b}}$$

$$(1.25) \quad \underline{\ddot{m}} = \underline{A} \underline{\ddot{b}}.$$

The following lemma from matrix theory is often useful.

Lemma 1.1. Given matrices \underline{B} ($k \times 1$), \underline{C} ($1 \times k$), \underline{D} ($k \times k$), and \underline{E} (1×1). Let

$$\underline{F} = \underline{D} + \underline{B} \underline{E} \underline{B}'.$$

Then, if the displayed inverses exist, we have

$$(1.26) \quad (\underline{I} + \underline{B} \underline{C})^{-1} = \underline{I} - \underline{B} (\underline{I} + \underline{C} \underline{B})^{-1} \underline{C}$$

$$(1.27) \quad (\underline{B}' \underline{F}^{-1} \underline{B})^{-1} \underline{B}' \underline{F}^{-1} = (\underline{B}' \underline{D}^{-1} \underline{B})^{-1} \underline{B}' \underline{D}^{-1}$$

$$(1.28) \quad \underline{B}' \underline{F}^{-1} \underline{B} = \underline{B}' \underline{D}^{-1} \underline{B} (\underline{I} + \underline{E} \underline{B}' \underline{D}^{-1} \underline{B})^{-1}.$$

(For proof, see De Vylder (1976a,p,139)).

Lemma 1.1, gives the following alternative expressions for $\hat{\underline{\beta}}$ and \underline{Z} if the displayed inverses exist:

$$(1.29) \quad \hat{\underline{\beta}} = (\underline{Y}' \underline{\Phi}^{-1} \underline{Y})^{-1} \underline{Y}' \underline{\Phi}^{-1} \underline{x} \quad (\text{from (1.27)})$$

$$(1.30) \quad \underline{Z} = \underline{\Lambda} \underline{Y}' \underline{\Phi}^{-1} \underline{Y} (\underline{I} + \underline{\Lambda} \underline{Y}' \underline{\Phi}^{-1} \underline{Y})^{-1} \quad (\text{from (1.28)}).$$

Jewell and Avenhaus (Jewell (1975a), Jewell & Avenhaus (1975)) have treated the case when the design matrix \underline{Y} is random. Jewell (1975a) has also treated the case when \underline{x} and \underline{m} are conditionally correlated given θ .

Example 1.1. Let x_j be the total claim amount of an insurance policy in the j -th insurance term it is running. We assume that the x_j 's are conditionally independent and identically distributed given an unknown random parameter θ , and that $EC(x_1|\theta)$ and $CE(x_1|\theta)$ exist and are non-zero and finite.

After r terms we seek \tilde{x}_{r+1} , the best linear inhomogeneous estimator of x_{r+1} (based on x_1, \dots, x_r) as a net premium for $(r+1)$ -th term. As x_1, \dots, x_r are exchangeable relative to x_{r+1} , \tilde{x}_{r+1} must be of the form

$$\gamma_0 + \gamma_1 \bar{x}_r,$$

where $\bar{x}_r = \frac{1}{r} \sum_{j=1}^r x_j$.

(1.6) and (1.7) give

$$\gamma_1 = \frac{C(x_{r+1}, \bar{x}_r)}{C(\bar{x}_r)}$$

$$\gamma_0 = (1 - \gamma_1) E(x_1).$$

We have

$$\gamma_1 = \frac{C(x_{r+1}, \bar{x}_r)}{C(\bar{x}_r)} =$$

$$\frac{EC(x_{r+1}, \bar{x}_r | \theta) + C(E(x_{r+1} | \theta), E(\bar{x}_r | \theta))}{EC(\bar{x}_r | \theta) + CE(\bar{x}_r | \theta)} =$$

$$\frac{\frac{CE(x_1 | \theta)}{\frac{1}{r} EC(x_1 | \theta) + CE(x_1 | \theta)}}{1} = \frac{r}{r + k},$$

where

$$k = \frac{EC(x_1 | \theta)}{CE(x_1 | \theta)}.$$

Hence

$$(1.31) \quad \tilde{x}_{r+1} = \frac{r}{r + k} \bar{x}_r + \frac{k}{r + k} E(x_1).$$

This now classical result was first shown by Bühlmann (1967), although a similar result had been shown by Bailey (1942, 1945).

We easily see by the strong law of large numbers that

$$\tilde{x}_{r+1} \xrightarrow{a.s.} E(x_1 | \theta)$$

as $r \rightarrow \infty$. This is a very satisfying result, as $E(x_1 | \theta)$ is the best possible estimator of x_{r+1} given x_1, \dots, x_r, θ .

We note the recursion

$$(1.32) \quad \tilde{x}_{r+1} = \frac{1}{r + \kappa} x_r + \frac{r - 1 + \kappa}{r + \kappa} \tilde{x}_r.$$

□

1H. We shall now briefly treat an interesting setup developed by Jewell (1975a,b).

In the model of subsection 1G let $\underline{x} = (\underline{x}_1', \underline{x}_2')'$ and $\underline{Y} = (\underline{Y}_1', \underline{Y}_2')'$. Here \underline{x}_i is a random $r_i \times 1$ -vector and \underline{Y}_i a non-random $r_i \times q$ -matrix for $i = 1, 2$. Both \underline{Y}_1 and \underline{Y}_2 are assumed to have rank q . \underline{x}_1 and \underline{x}_2 are assumed independent given θ . We have

$$E(\underline{x}_i | \theta) = \underline{Y}_i \underline{b}(\theta)$$

and assume that

$$\underline{\Phi}_i = EC(\underline{x}_i | \theta)$$

is invertible for $i = 1, 2$.

From (1.22), (1.29), and (1.30) we get the best linear inhomogeneous estimator of $\underline{b}(\theta)$ based on \underline{x}_1

$$\tilde{\underline{b}}_1 = \underline{Z}_1 \hat{\underline{b}}_1 + (\underline{I} - \underline{Z}_1) \underline{\beta}$$

with

$$\hat{\underline{b}}_1 = (\underline{Y}_1' \underline{\Phi}_1^{-1} \underline{Y}_1)^{-1} \underline{Y}_1' \underline{\Phi}_1^{-1} \underline{x}_1$$

and

$$\underline{Z}_1 = \underline{A} \underline{Y}_1' \underline{\Phi}_1^{-1} \underline{Y}_1 (\underline{I} + \underline{A} \underline{Y}_1' \underline{\Phi}_1^{-1} \underline{Y}_1)^{-1}.$$

Following Jewell we define the preposterior covariance of the parameter estimation error of $\underline{\tilde{b}}_1$

$$\underline{\underline{\Lambda}}_1 = C(\underline{\underline{b}}(\theta) - \underline{\underline{\beta}}_1).$$

After some trivial calculus we get

$$\begin{aligned} \underline{\underline{\Lambda}}_1 &= (\underline{\underline{I}} - \underline{\underline{Z}}_1) \underline{\underline{\Lambda}} = \\ &(\underline{\underline{\Lambda}}^{-1} + \underline{\underline{Y}}_1' \underline{\underline{\Phi}}_1^{-1} \underline{\underline{Y}}_1)^{-1}. \end{aligned}$$

Jewell has shown that the best linear inhomogeneous estimator of $\underline{\underline{b}}(\theta)$ based on $\underline{\underline{x}}_1$ and $\underline{\underline{x}}_2$ can be written

$$\underline{\underline{\beta}}_2 = \underline{\underline{Z}}_2 \hat{\underline{\underline{b}}}_2 + (\underline{\underline{I}} - \underline{\underline{Z}}_2) \underline{\underline{\beta}}_1$$

with

$$\hat{\underline{\underline{b}}}_2 = (\underline{\underline{Y}}_2' \underline{\underline{\Phi}}_2^{-1} \underline{\underline{Y}}_2)^{-1} \underline{\underline{Y}}_2' \underline{\underline{\Phi}}_2^{-1} \underline{\underline{x}}_2$$

and

$$\underline{\underline{Z}}_2 = \underline{\underline{\Lambda}}_1 \underline{\underline{Y}}_2' \underline{\underline{\Phi}}_2^{-1} \underline{\underline{Y}}_2 (\underline{\underline{I}} + \underline{\underline{\Lambda}}_1 \underline{\underline{Y}}_2' \underline{\underline{\Phi}}_2^{-1} \underline{\underline{Y}}_2)^{-1}.$$

Then we see that the best linear inhomogeneous estimator of $\underline{\underline{b}}(\theta)$ based on $\underline{\underline{x}}_1$ and $\underline{\underline{x}}_2$ has the same form as the one based on $\underline{\underline{x}}_2$ with $\underline{\underline{\beta}}$ replaced by $\underline{\underline{\beta}}_1$ and $\underline{\underline{\Lambda}}$ replaced by $\underline{\underline{\Lambda}}_1$. This gives us a good method of updating our estimates when we get more data.

The preposterior covariance of the parameter estimation error of $\underline{\underline{\beta}}_2$ is

$$\begin{aligned} \underline{\underline{\Lambda}}_2 &= (\underline{\underline{\Lambda}}^{-1} + \underline{\underline{Y}}' \underline{\underline{\Phi}}^{-1} \underline{\underline{Y}})^{-1} = \\ &(\underline{\underline{\Lambda}}^{-1} + \underline{\underline{Y}}_1' \underline{\underline{\Phi}}_1^{-1} \underline{\underline{Y}}_1 + \underline{\underline{Y}}_2' \underline{\underline{\Phi}}_2^{-1} \underline{\underline{Y}}_2)^{-1} \end{aligned}$$

giving

$$\begin{aligned} \underline{\underline{\Lambda}}_2 &= (\underline{\underline{\Lambda}}_1^{-1} + \underline{\underline{Y}}_2' \underline{\underline{\Phi}}_2^{-1} \underline{\underline{Y}}_2)^{-1} = \\ &(\underline{\underline{I}} - \underline{\underline{Z}}_2) \underline{\underline{\Lambda}}_1 \end{aligned}$$

Thus the preposterior covariance of the parameter estimation error can be updated in the same way as the estimates.

For further details we refer to Jewell (1975a,b).

11. In the model of Example 1.1 it is natural to ask if we should use data from other policies in our credibility premium. This can be done in two ways:

- i) by making optimal linear estimators, where data from the other policies appear linearly.
- ii) by using data from other policies to estimate unknown parameters (e.g. κ and $E(x_1)$ in Example 1.1).

In the present thesis we are mainly to concentrate on case i), although case ii) will be touched in section 2. Case ii) has been studied by Bühlmann & Strauß (1970), Hachemeister (1975), Norberg (1978), and De Vylder (1978a,b).

As a start, in the model of subsection 1F let $\underline{x} = (\underline{x}_1', \underline{x}_2')'$ and $\underline{y} = (\underline{y}_1', \underline{y}_2')'$. Here \underline{x}_i is a random $r_i \times 1$ -vector and \underline{y}_i a non-random $r_i \times q$ -matrix for $i = 1, 2$ ($r_1 + r_2 = r$). \underline{y}_1 has rank q . (This implies that \underline{y} has rank q .) \underline{x}_2 and θ are independent. We assume that

$$E(\underline{x}_1 | \theta) = \underline{y}_1 \underline{b}(\theta)$$

$$EC(\underline{x}_1 | \theta) = \underline{\phi},$$

and let

$$\underline{\Sigma}_i = C(\underline{x}_i) \quad i = 1, 2.$$

Comparing with Example 1.1, \underline{x}_1 could be data from the policy we want to tariff, and \underline{x}_2 data from other policies.

We get

$$(1.33) \quad C(\underline{m}, \underline{x}') = \underline{A} \underline{\Lambda} (\underline{y}_1', \underline{0}).$$

As $(\underline{y}_1', \underline{0})$ and \underline{y} have rank q and $\underline{\Sigma}^{-1}$ rank r , $(\underline{y}_1', \underline{0}) \underline{\Sigma}^{-1} \underline{y}$ must have full rank and is thus invertible. Then (1.13), (1.14), and (1.33) give

$$(1.34) \quad \underline{\tilde{m}} = \underline{A} [\underline{Z} \hat{\underline{b}} + (\underline{I} - \underline{Z}) \underline{\beta}]$$

$$(1.35) \quad \underline{\ddot{m}} = \underline{A} [\underline{Z} \hat{\underline{b}} + (\underline{I} - \underline{Z}) \hat{\underline{\beta}}],$$

where

$$\underline{Z} = \underline{A} (\underline{Y}_1', \underline{0}) \underline{\Sigma}^{-1} \underline{Y}$$

and

$$(1.36) \quad \hat{\underline{b}} = ((\underline{Y}_1', \underline{0}) \underline{\Sigma}^{-1} \underline{Y})^{-1} (\underline{Y}_1', \underline{0}) \underline{\Sigma}^{-1} \underline{x}.$$

Corresponding to (1.22) and (1.23) we get

$$\underline{\tilde{b}} = \underline{Z} \hat{\underline{b}} + (\underline{I} - \underline{Z}) \underline{\beta}$$

$$\underline{\ddot{b}} = \underline{Z} \hat{\underline{b}} + (\underline{I} - \underline{Z}) \hat{\underline{\beta}},$$

and (1.24) and (1.25) are still valid.

1J. We now make the further assumption that \underline{x}_1 and \underline{x}_2 are independent. Then

$$\underline{\Sigma} = \begin{pmatrix} \underline{\Sigma}_1 & \underline{0} \\ \underline{0} & \underline{\Sigma}_2 \end{pmatrix},$$

where $\underline{\Sigma}_i$ denotes the covariance matrix of \underline{x}_i . We note that $\underline{\Sigma}$ is invertible if and only if both $\underline{\Sigma}_1$ and $\underline{\Sigma}_2$ are invertible.

We now get

$$(1.37) \quad \underline{Z} = \underline{A} \underline{Y}_1' \underline{\Sigma}_1^{-1} \underline{Y}_1$$

$$(1.38) \quad \hat{\underline{b}} = (\underline{Y}_1' \underline{\Sigma}_1^{-1} \underline{Y}_1)^{-1} \underline{Y}_1 \underline{\Sigma}_1^{-1} \underline{x}_1.$$

1K. To distinguish from the unbiasedness concept defined in subsection 1D we shall say that an estimator \dot{m} of m is θ -unbiased if $E(\dot{m}|\theta) = E(m|\theta)$.

We shall assume the same model as in the previous subsection.

For a linear θ -unbiased estimator

$$\dot{m} = g_0 + \underline{g}_1' \underline{x}_1 + \underline{g}_2' \underline{x}_2$$

we must have

$$\underline{a}'\underline{b}(\theta) = g_0 + \underline{g}_1' \underline{Y}_1 \underline{b}(\theta) + \underline{g}_2' \underline{Y}_2 \underline{\beta}$$

or equivalently

$$(1.39) \quad (\underline{a}' - \underline{g}_1' \underline{Y}_1) \underline{b}(\theta) = g_0 + \underline{g}_2' \underline{Y}_2 \underline{\beta}.$$

From this follows that $(\underline{a}' - \underline{g}_1' \underline{Y}_1) \underline{b}(\theta)$ must have variance zero, that is,

$$(\underline{a}' - \underline{g}_1' \underline{Y}_1)' \underline{\Lambda} (\underline{a}' - \underline{g}_1' \underline{Y}_1) = 0.$$

As $\underline{\Lambda}$ was assumed positive definite, this implies that

$$(1.40) \quad \underline{g}_1' \underline{Y}_1 = \underline{a}'.$$

Inserting (1.40) in (1.39) gives

$$g_0 + \underline{g}_2' \underline{Y}_2 \underline{\beta} = 0.$$

We have

$$E(\hat{m} - m)^2 = E(\underline{g}_1' \underline{x}_1 - m)^2 + E(g_0 + \underline{g}_2' \underline{x}_2)^2 \geq E(\underline{g}_1' \underline{x}_1 - m)^2.$$

As $\underline{g}_1' \underline{x}_1$ is also a θ -unbiased linear estimator of m , the best linear θ -unbiased estimator of m must be of the form $\underline{g}_1' \underline{x}_1$. But then $\hat{m}^{(\theta)}$, the best linear θ -unbiased estimator of m , must be the best linear unbiased homogeneous estimator of m based on \underline{x}_1 , and (1.21) gives

$$(1.41) \quad \hat{m}^{(\theta)} = \underline{a}' \hat{\underline{b}}.$$

The vector generalization

$$\hat{\underline{m}}^{(\theta)} = \underline{A} \hat{\underline{b}}$$

of (1.41) is obvious. In particular we observe that $\hat{\underline{b}}$ is the best linear θ -unbiased estimator of $\underline{b}(\theta)$.

We now see that the optimal linear estimator of $\underline{b}(\theta)$ is a weighted average of the best linear θ -unbiased estimator of $\underline{b}(\theta)$ and $E(\underline{b}(\theta))$ (inhomogeneous case), or the best linear θ -unbiased estimator of $\underline{b}(\theta)$ and the best linear unbiased estimator of $E(\underline{b}(\theta))$ (unbiased homogeneous case). The weights are the same

in both cases. The optimal estimator of m is obtained by multiplying \underline{A} by the optimal estimator of $\underline{b}(\theta)$.

In the model of subsection 1G, $\hat{\underline{\beta}}$ is both the best linear unbiased estimator of $\underline{\beta}$ and the best linear θ -unbiased estimator of $\underline{b}(\theta)$. Hence we get the simple expression

$$\hat{\underline{m}} = \underline{A} \hat{\underline{\beta}}.$$

$\hat{\underline{b}}$ is not generally θ -unbiased in the model of subsection 1I. This will be further discussed in subsection 2B.

1L. Let $(\underline{x}'_1, \underline{m}'_1, \theta_1), \dots, (\underline{x}'_n, \underline{m}'_n, \theta_n)$ be independent random vectors. $\theta_1, \dots, \theta_n$ are unknown and identically distributed.

For each i \underline{x}_i is an observable random $r_i \times 1$ -vector and \underline{m}_i an unknown random $s_i \times 1$ -vector. \underline{x}_i and \underline{m}_i are independent given θ_i .

$$E(\underline{x}_i | \theta_i) = \underline{Y}_i \underline{b}(\theta_i),$$

$$E(\underline{m}_i | \theta_i) = \underline{A}_i \underline{b}(\theta_i).$$

Here $\underline{b}(\theta_i)$ is a $q \times 1$ vector function of θ_i ; \underline{Y}_i is a non-random $r_i \times q$ -matrix of rank q , and \underline{A}_i a non-random $s_i \times q$ -matrix.

We assume that

$$\underline{A} = C(\underline{b}(\theta_1))$$

and

$$\underline{\Sigma}_i = C(\underline{x}_i) \quad i = 1, \dots, n$$

are invertible.

This model was first studied by Hachemeister (1975).

Let

$$\underline{\beta} = E(\underline{b}(\theta_1)),$$

$$\underline{Y} = (\underline{Y}'_1, \dots, \underline{Y}'_n)',$$

$$\underline{X} = (\underline{X}'_1, \dots, \underline{X}'_n)',$$

and

$$\underline{\Sigma} = C(\underline{X}) = \begin{pmatrix} \underline{\Sigma}_1 & \underline{0} & \dots & \underline{0} \\ \underline{0} & \underline{\Sigma}_2 & \dots & \underline{0} \\ \vdots & \vdots & & \vdots \\ \underline{0} & \underline{0} & \dots & \underline{\Sigma}_n \end{pmatrix}.$$

(1.11), (1.34), (1.35), (1.37), and (1.38) now give the optimal estimators of \underline{m}_i :

$$(1.42) \quad \tilde{\underline{m}}_i = \underline{A}_i [\underline{Z}_i \hat{\underline{b}}_i + (\underline{I} - \underline{Z}_i) \underline{\beta}]$$

and

$$\ddot{\underline{m}}_i = \underline{A}_i [\underline{Z}_i \hat{\underline{b}}_i + (\underline{I} - \underline{Z}_i) \hat{\underline{\beta}}],$$

where

$$\underline{Z}_i = \underline{A} \underline{Y}'_i \underline{\Sigma}_i^{-1} \underline{Y}_i,$$

$$\hat{\underline{b}}_i = (\underline{Y}'_i \underline{\Sigma}_i^{-1} \underline{Y}_i)^{-1} \underline{Y}'_i \underline{\Sigma}_i^{-1} \underline{X}_i,$$

and

$$\hat{\underline{\beta}} = (\underline{Y}' \underline{\Sigma}^{-1} \underline{Y})^{-1} \underline{Y}' \underline{\Sigma}^{-1} \underline{X}.$$

We have

$$\begin{aligned} \hat{\underline{\beta}} &= (\underline{Y}' \underline{\Sigma}^{-1} \underline{Y})^{-1} \underline{Y}' \underline{\Sigma}^{-1} \underline{X} = \left(\sum_{j=1}^n \underline{Y}'_j \underline{\Sigma}_j^{-1} \underline{Y}_j \right)^{-1} \sum_{j=1}^n \underline{Y}'_j \underline{\Sigma}_j^{-1} \underline{X}_j = \\ &= \left(\sum_{j=1}^n \underline{A} \underline{Y}'_j \underline{\Sigma}_j^{-1} \underline{Y}_j \right)^{-1} \sum_{j=1}^n \underline{A} \underline{Y}'_j \underline{\Sigma}_j^{-1} \underline{Y}_j (\underline{Y}'_j \underline{\Sigma}_j^{-1} \underline{Y}_j)^{-1} \underline{Y}_j \underline{\Sigma}_j^{-1} \underline{X}_j = \\ &= \left(\sum_{j=1}^n \underline{Z}_j \right)^{-1} \sum_{j=1}^n \underline{Z}_j \hat{\underline{b}}_j. \end{aligned}$$

We then get

$$(1.43) \quad \hat{\underline{m}}_i = \underline{A}_i [\underline{Z}_i \hat{\underline{b}}_i + (\underline{I} - \underline{Z}_i) (\sum_{j=1}^n \underline{Z}_j)^{-1} \sum_{j=1}^n \underline{Z}_j \hat{\underline{b}}_j].$$

This result was first shown by Taylor (1977).

In the special case when $\underline{x}_1, \dots, \underline{x}_n$ are identically distributed, we get

$$\hat{\underline{g}} = \frac{1}{n} \sum_{j=1}^n \hat{\underline{b}}_j = \bar{\underline{b}}$$

and

$$\hat{\underline{m}}_i = \underline{A}_i [\underline{Z}_i \hat{\underline{b}}_i + (\underline{I} - \underline{Z}_i) \bar{\underline{b}}].$$

In this case $\bar{\underline{b}} \xrightarrow{a.s.} \underline{g}$ as $n \rightarrow \infty$ by the strong law of large numbers.

Example 1.2. Assume that we have an insurance portfolio consisting of n independent policies of the type described in Example 1.1. x_{ij} is the total claim amount of the i -th policy in the j -th term it is running. x_{i1}, x_{i2}, \dots are conditionally independent and identically distributed given an unknown random parameter θ_i . i -th policy has been running in r_i terms. We assume that $\theta_1, \dots, \theta_n$ are independent and identically distributed, and that the conditional cumulative distribution function of x_{ij} given θ_i is on the form $F(\cdot | \theta_i)$ with F independent of i .

We want to estimate x_{k, r_k+1} with linear estimators based on the observed claim amounts in the portfolio. From Example 1.1 and (1.42) we get the best linear inhomogeneous estimator

$$(1.44) \quad \tilde{x}_{k, r_k+1} = \frac{r_k}{r_k + \kappa} \bar{x}_{k, r_k} + \frac{\kappa}{r_k + \kappa} E(x_{11}),$$

where

$$\kappa = \frac{EC(x_{11}|\theta_1)}{CE(x_{11}|\theta_1)}$$

is assumed to exist, and

$$\bar{x}_{k,r_k} = \frac{1}{r_k} \sum_{j=1}^{r_k} x_{kj}.$$

From (1.42), (1.43), and (1.44) we now easily get

$$(1.45) \quad \ddot{x}_{k,r_k+1} = \frac{r_k}{r_k + \kappa} \bar{x}_{k,r_k} + \frac{\kappa}{r_k + \kappa} \frac{\sum_{i=1}^n \frac{r_i}{r_i + \kappa} \bar{x}_{i,r_i}}{\sum_{i=1}^n \frac{r_i}{r_i + \kappa}}.$$

In the special case $r_1 = \dots = r_n = r$, (1.45) takes the simple form

$$\ddot{x}_{k,r+1} = \frac{r}{r + \kappa} \bar{x}_{kr} + \frac{\kappa}{r + \kappa} \bar{\bar{x}}_r$$

where

$$\bar{\bar{x}}_r = \frac{1}{n} \sum_{i=1}^n \bar{x}_{ir}.$$

□

1M. Let θ be an unknown random variable. We want to estimate a random variable m and assume that $E(m|\theta) = \underline{a}'\underline{b}(\theta)$, where \underline{a} is a non-random $q \times 1$ -vector and $\underline{b}(\theta)$ a $q \times 1$ vector function of θ . Then, as we have seen, our estimator would often be on the form

$$(1.46) \quad m^* = \underline{a}'[\underline{Z} \underline{b}^* + (\underline{I} - \underline{Z}) \underline{\beta}],$$

where $\underline{\beta} = E(\underline{b}(\theta))$, \underline{b}^* is a θ -unbiased estimator of $\underline{b}(\theta)$, and \underline{Z} is a non-random $q \times q$ -matrix.

Suppose now that after we have computed the matrix \underline{Z} (which may have been quite a job), we find in our files an old estimate $\dot{\underline{b}}$ of $\underline{b}(\theta)$ based on some other data. As we do not feel too happy with the thought of a lot of work to find an optimal estimator of m based on both $\dot{\underline{b}}$ and our recent data, we feel tempted to simply replace $\underline{\beta}$ in (1.46) by $\dot{\underline{b}}$ to give an estimator

$$\dot{m} = \underline{a}'[\underline{Z} \dot{\underline{b}}^* + (\underline{I} - \underline{Z}) \dot{\underline{b}}].$$

Our natural question is then: Is \dot{m} a better estimator of m than m^* , that is, is

$$E(\dot{m} - m)^2 < E(m^* - m)^2 ?$$

The following theorem gives a partial answer.

Theorem 1.3. Suppose that $\dot{\underline{b}}$ is independent of m and $\dot{\underline{b}}^*$ given θ . Then $E(\dot{m} - m)^2 \leq E(m^* - m)^2$ if and only if

$$E(\underline{a}'(\underline{I} - \underline{Z})(\dot{\underline{b}} - \underline{b}(\theta)))^2 \leq C(\underline{a}'(\underline{I} - \underline{Z}) \underline{b}(\theta)),$$

and $E(\dot{m} - m)^2 < E(m^* - m)^2$ if and only if

$$E(\underline{a}'(\underline{I} - \underline{Z})(\dot{\underline{b}} - \underline{b}(\theta)))^2 < C(\underline{a}'(\underline{I} - \underline{Z}) \underline{b}(\theta)).$$

Proof. We have

$$E(\dot{m} - m)^2 = E(\underline{a}'[\underline{Z} \dot{\underline{b}}^* + (\underline{I} - \underline{Z}) \dot{\underline{b}}] - m)^2 =$$

$$EE(\{\underline{a}'(\underline{I} - \underline{Z})(\dot{\underline{b}} - \underline{b}(\theta)) + \underline{a}'[\underline{Z} \dot{\underline{b}}^* + (\underline{I} - \underline{Z}) \underline{b}(\theta)] - m\}^2 | \theta).$$

As $\underline{a}'(\underline{I} - \underline{Z})(\dot{\underline{b}} - \underline{b}(\theta))$ and $\underline{a}'[\underline{Z} \dot{\underline{b}}^* + (\underline{I} - \underline{Z}) \underline{b}(\theta)] - m$ are independent given θ and

$$E(\underline{a}'[\underline{Z} \dot{\underline{b}}^* + (\underline{I} - \underline{Z}) \underline{b}(\theta)] - m | \theta) = 0,$$

it follows that

$$(1.47) \quad E(\dot{m} - m)^2 = E(\underline{a}'(\underline{I} - \underline{Z})(\dot{\underline{b}} - \underline{b}(\theta)))^2 +$$

$$E(\underline{a}'[\underline{Z} \dot{\underline{b}}^* + (\underline{I} - \underline{Z}) \underline{b}(\theta)] - m)^2.$$

Replacing \dot{m} with m^* and $\dot{\underline{b}}$ with $\underline{\beta}$ above gives

$$(1.48) \ E(m^* - m)^2 = C(\underline{a}'(\underline{I} - \underline{Z}) \underline{b}(\theta)) + \\ E(\underline{a}'[\underline{Z} \underline{b}^* + (\underline{I} - \underline{Z}) \underline{b}(\theta)] - m)^2$$

The theorem now follows from (1.47) and (1.48).

Q E D.

Definition 1.2. Let \underline{B} and \underline{C} be two $n \times n$ -matrices. We shall say that \underline{C} is less than \underline{B} if $\underline{B} - \underline{C}$ is non-negative definite and $\underline{B} \neq \underline{C}$.

We can now state the following corollary to Theorem 1.3.

Corollary 1.1. If $\dot{\underline{b}}$ is independent of \underline{b}^* and m given θ , and $E((\dot{\underline{b}} - \underline{b}(\theta))(\dot{\underline{b}} - \underline{b}(\theta))')$ is less than $C(\underline{b}(\theta))$, then $E(\dot{\underline{m}} - m)^2 \leq E(m^* - m)^2$.

Proof: $E((\dot{\underline{b}} - \underline{b}(\theta))(\dot{\underline{b}} - \underline{b}(\theta))')$ is less than $C(\underline{b}(\theta))$ implies that $\underline{a}'(\underline{I} - \underline{Z}) E((\dot{\underline{b}} - \underline{b}(\theta))(\dot{\underline{b}} - \underline{b}(\theta))')(\underline{I} - \underline{Z})' \underline{a} \leq \underline{a}'(\underline{I} - \underline{Z}) C(\underline{b}(\theta))(\underline{I} - \underline{Z})' \underline{a}$,

which is equivalent with

$$E(\underline{a}'(\underline{I} - \underline{Z}) (\dot{\underline{b}} - \underline{b}(\theta))^2 \leq C(\underline{a}'(\underline{I} - \underline{Z}) \underline{b}(\theta)),$$

And the corollary follows from Theorem 1.3.

Q E D.

It may be that we have to use m even if we wanted to make a better estimator based on both $\dot{\underline{b}}$ and our recent data, because we do not have sufficient knowledge about the properties of $\dot{\underline{b}}$. For instance, if we have the model of subsection 1H, we may not know the preposterior covariance of the parameter estimation error of $\dot{\underline{b}}$.

Let us look at the special case with $q = 1$ and

$E(m|\theta) = \underline{b}(\theta) = b(\theta)$. Let $\underline{\beta} = \beta$, $\underline{b}^* = b^*$, $\dot{\underline{b}} = \dot{b}$, and $\underline{Z} = \zeta$.

Then

$$m^* = \zeta b^* + (1 - \zeta) \beta$$

and

$$\dot{m} = \zeta \dot{b}^* + (1 - \zeta) \dot{b}.$$

Theorem 1.3 now reduces to

Corollary 1.2. Suppose that \dot{b} is independent of m and b^* given θ . Then $E(\dot{m} - m)^2 \leq E(m^* - m)^2$ if and only if

$$E(\dot{b} - b(\theta))^2 \leq E(\beta - b(\theta))^2,$$

and $E(\dot{m} - m)^2 < E(m^* - m)^2$ if and only if

$$(1.49) \quad E(\dot{b} - b(\theta))^2 < E(\beta - b(\theta))^2,$$

From (1.17) follows that (1.49) is equivalent with

$$E(\dot{b} - m)^2 < E(\beta - m)^2.$$

This makes sense! We ought to choose the one of \dot{b} and β that lies closest (in least mean squares sense) to the quantity we want to estimate.

Example 1.3. Let x_j be the total claim amount of an insurance policy in the j -th insurance term it is running. We assume that the x_j 's are conditionally independent and identically distributed given an unknown random parameter θ . We furthermore assume that $\kappa = \frac{EC(x_1|\theta)}{CE(x_1|\theta)}$ exists and that $0 < \kappa < \infty$.

From (1.31) we find the Best linear inhomogeneous estimator of x_{r+1} based on x_r

$$x_{r+1}^* = \frac{1}{1 + \kappa} x_r + \frac{\kappa}{1 + \kappa} E(x_1).$$

Now let

$$(1.50) \quad \begin{aligned} \dot{x}_2 &= x_2^* \\ \dot{x}_{r+1} &= \frac{1}{1+\kappa} x_r + \frac{\kappa}{1+\kappa} \dot{x}_r \end{aligned} \quad r = 2, 3, \dots$$

From subsection 1 F follows that an estimator of x_{r+1} is optimal if and only if it is an optimal estimator of $E(x_1|\theta)$.

Since \dot{x}_2 is the best linear inhomogeneous estimator of $E(x_1|\theta)$ based on x_1 , \dot{x}_2 must particularly be better than $E(x_1)$, that is, $E(\dot{x}_2 - E(x_1|\theta))^2 < E(E(x_1) - E(x_1|\theta))^2$. It now follows from Corollary 1.2 that \dot{x}_3 is a better estimator of $E(x_1|\theta)$ than x_3^* , which is again a better estimator of $E(x_1|\theta)$ than $E(x_1)$.

This gives

$$E(\dot{x}_3 - E(x_1|\theta))^2 < E(x_3^* - E(x_1|\theta))^2 < E(E(x_1) - E(x_1|\theta))^2.$$

Suppose now that

$$E(\dot{x}_r - E(x_1|\theta))^2 < E(x_r^* - E(x_1|\theta))^2 < E(E(x_1) - E(x_1|\theta))^2.$$

As above it follows that

$$E(\dot{x}_{r+1} - E(x_1|\theta))^2 < E(x_{r+1}^* - E(x_1|\theta))^2 < E(E(x_1) - E(x_1|\theta))^2,$$

and we have thereby proved by induction that \dot{x}_{r+1} is a better estimator of $E(x_1|\theta)$ than x_{r+1}^* for $r = 2, 3, \dots$, or equivalently that \dot{x}_{r+1} is a better estimator of x_{r+1} than x_{r+1}^* for $r = 2, 3, \dots$.

From (1.31) we have that the best linear inhomogeneous estimator of x_{r+1} based on x_1, \dots, x_r is

$$\tilde{x}_{r+1} = \frac{r}{r+\kappa} \bar{x}_r + \frac{\kappa}{r+\kappa} E(x_1).$$

We now have

$$E(\tilde{x}_{r+1} - x_{r+1})^2 < E(\dot{x}_{r+1} - x_{r+1})^2 < E(x_{r+1}^* - x_{r+1})^2 < E(E(x_1) - x_{r+1})^2,$$

giving a rank ordering of \tilde{x}_{r+1} , \dot{x}_{r+1} , x_{r+1}^* , and $E(x_1)$ as estimators of x_{r+1} .

From (1.50) follows that

$$\dot{x}_{r+1} = \frac{1}{1+\kappa} \sum_{j=1}^r \left(\frac{\kappa}{1+\kappa} \right)^{r-j} x_j + \left(\frac{\kappa}{1+\kappa} \right)^r E(x_1).$$

We see that the weights given to the observations are geometric. Credibility estimators with geometric weights have been studied by Gerber & Jones (1973,1975).

□

2. An insurance model with collective seasonal random factors.

A model for an insurance portfolio is developed in which each insurance term is characterized by an unknown random parameter influencing the claim amounts of all the policies in the portfolio in that term. Two different ratemaking procedures are developed, and an approach by Welten (1968) is briefly summarized.

2A. Assume that we have an insurance portfolio consisting of n policies that have been running in the same r insurance terms. x_{ij} is the total claim amount of i -th policy in j -th term. x_{i1}, x_{i2}, \dots are conditionally independent and identically distributed given an unknown random parameter θ_i , and the x_{ij} 's are independent of θ_k for $k \neq i$. We assume that $\theta_1, \dots, \theta_n$ are independent and identically distributed.

In Example 1.2 we assumed that the claim amounts x_{1j}, \dots, x_{nj} from the policies in a term j were independent. However, this is not always the case in practice. In many situations there seem to be seasonal random factors influencing the claim amounts of the whole portfolio. We shall look at some specific examples:

Example 2.1. Motor insurance. Suppose that one winter the weather has given extremely icy roads. This could lead to many car accidents.

Example 2.2. Marine insurance. In a stormy year there could be lots of shipwrecks.

Example 2.3. Forest fire insurance. A dry summer could lead to many forest-fires.

Example 2.4. Burglary insurance. The electricity cut in New York in 1977 caused an enormous amount of burglaries.

□

In these examples it seems natural to assume that to each insurance term j there is connected an unknown random parameter η_j , and that

- i) x_{1j}, \dots, x_{nj} are independent given η_j ,
- ii) x_{1j}, \dots, x_{nj} are independent of η_l for $l \neq j$,
- iii) η_1, η_2, \dots are independent and identically distributed,
- iv) the η_j 's are independent of the θ_i 's,
- v) x_{ij} and x_{kl} are independent if both $i \neq k$ and $j \neq l$.

We also assume that the conditional cumulative distribution of x_{ij} given θ_i and η_j is on the form $F(\cdot | \theta_i, \eta_j)$ with F independent of i and j .

Let

$$b(\theta_1) = E(x_{11} | \theta_1)$$

$$c(\eta_1) = E(x_{11} | \eta_1)$$

$$\beta = E(x_{11})$$

$$\bar{x}_i = \frac{1}{r} \sum_{j=1}^r x_{ij} \quad i = 1, \dots, n$$

$$\bar{\bar{x}} = \frac{1}{n} \sum_{i=1}^n \bar{x}_i.$$

The necessary properties of occurring second order moments are silently assumed.

2B. We want an expression for $\tilde{x}_{k,r+1}$, the best linear inhomogeneous estimator of $x_{k,r+1}$.

As $(x_{11}, \dots, x_{n1})'$, \dots , $(x_{1r}, \dots, x_{nr})'$ are exchangeable relative to $x_{k,r+1}$, the formula will depend on only $\bar{x}_1, \dots, \bar{x}_n$. Furthermore, the \bar{x}_i 's for $i \neq k$ are exchangeable relative to $x_{k,r+1}$, and then we can write

$$(2.1) \quad \tilde{x}_{k,r+1} = \gamma_0 + \gamma_1 \bar{x}_k + \gamma_2 \bar{\bar{x}}.$$

From (1.1) and (1.3) we get

$$(2.2) \quad \gamma_0 = (1 - \gamma_1 - \gamma_2) \beta$$

$$(2.3) \quad \gamma_1 C(\bar{x}_k) + \gamma_2 C(\bar{\bar{x}}, \bar{x}_k) = C(x_{k,r+1}, \bar{x}_k)$$

$$(2.4) \quad \gamma_1 C(\bar{x}_k, \bar{\bar{x}}) + \gamma_2 C(\bar{\bar{x}}) = C(x_{k,r+1}, \bar{\bar{x}}).$$

The following relations are easily verified:

$$(2.5) \quad C(\bar{x}_k, \bar{\bar{x}}) = C(\bar{\bar{x}})$$

$$(2.6) \quad C(x_{k,r+1}, \bar{x}_k) = C(b(\theta_1))$$

$$(2.7) \quad C(x_{k,r+1}, \bar{\bar{x}}) = \frac{1}{n} C(b(\theta_1))$$

$$C(\bar{\bar{x}}) = \frac{1}{n} C(\bar{x}_1) + (1 - \frac{1}{n}) C(\bar{x}_1, \bar{x}_2)$$

$$C(\bar{x}_1) = C(b(\theta_1)) + \frac{1}{r} EC(x_{11} | \theta_1)$$

$$C(\bar{x}_1, \bar{x}_2) = \frac{1}{r} EC(x_{11}, x_{21} | \theta_1, \theta_2) = \frac{1}{r} C(c(\eta_1)).$$

Using (2.5), (2.6), and (2.7), (2.3) and (2.4) can now be rewritten

$$(2.8) \quad \gamma_1 C(\bar{x}_1) + \gamma_2 C(\bar{\bar{x}}) = C(b(\theta_1))$$

$$(2.9) \quad (\gamma_1 + \gamma_2) C(\bar{\bar{x}}) = \frac{1}{n} C(b(\theta_1)).$$

Subtracting (2.9) from (2.8) gives

$$\gamma_1 (C(\bar{x}_1) - C(\bar{x})) = (1 - \frac{1}{n}) C(b(\theta_1)).$$

$$\gamma_1 = (1 - \frac{1}{n}) \frac{C(b(\theta_1))}{C(\bar{x}_1) - C(\bar{x})} =$$

$$(1 - \frac{1}{n}) \frac{C(b(\theta_1))}{C(\bar{x}_1) - \frac{1}{n} C(\bar{x}_1) - (1 - \frac{1}{n}) C(\bar{x}_1, \bar{x}_2)} =$$

$$\frac{C(b(\theta_1))}{C(\bar{x}_1) - C(\bar{x}_1, \bar{x}_2)} = \frac{C(b(\theta_1))}{C(b(\theta_1)) + \frac{1}{r} EC(x_{11}|\theta_1) - \frac{1}{r} C(c(\eta_1))}$$

From (2.9) we get

$$\gamma_1 + \gamma_2 = \frac{1}{n} \frac{C(b(\theta_1))}{C(\bar{x})} = \frac{C(b(\theta_1))}{C(\bar{x}_1) + (n-1) C(\bar{x}_1, \bar{x}_2)} =$$

$$\frac{C(b(\theta_1))}{C(b(\theta_1)) + \frac{1}{r} EC(x_{11}|\theta_1) + (n-1) \frac{1}{r} C(c(\eta_1))}$$

Letting

$$(2.10) \quad \kappa = \frac{EC(x_{11}|\theta_1)}{C(b(\theta_1))}$$

and

$$(2.11) \quad \rho = \frac{C(c(\eta_1))}{C(b(\theta_1))}$$

gives

$$(2.12) \quad \gamma_1 = \frac{r}{r + \kappa - \rho}$$

and

$$(2.13) \quad \gamma_1 + \gamma_2 = \frac{r}{r + \kappa + (n-1)\rho}.$$

From (2.1), (2.2), (2.12), and (2.13) we now get

$$(2.14) \quad \tilde{x}_{k,r+1} = \frac{r}{r + \kappa - \rho} (\bar{x}_k - \bar{\bar{x}}) + \frac{r}{r + \kappa + (n-1)\rho} \bar{\bar{x}} + \frac{\kappa + (n-1)\rho}{r + \kappa + (n-1)\rho} \beta.$$

Since $x_{k,r+1}$ is independent of x_{ij} for $i = 1, \dots, n$ and $j = 1, \dots, r$ given θ_k , and θ_k is independent of the x_{ij} 's for $i \neq k$ and $j = 1, \dots, r$, we have the model of subsection 1I. Rewriting (2.14) gives

$$\tilde{x}_{k,r+1} = \frac{r}{r + \kappa + (n-1)\rho} \left[\frac{r + \kappa + (n-1)\rho}{r + \kappa - \rho} \bar{x}_k - \frac{n\rho}{r + \kappa - \rho} \bar{\bar{x}} \right] + \frac{\kappa + (n-1)\rho}{r + \kappa + (n-1)\rho} \beta$$

corresponding to (1.34), and we get

$$\hat{b}_k = \frac{r + \kappa + (n-1)\rho}{r + \kappa - \rho} \bar{x}_k - \frac{n\rho}{r + \kappa - \rho} \bar{\bar{x}}$$

corresponding to \hat{b} defined by (1.36).

We have

$$E(\hat{b}_k | \theta_k) = \frac{r + \kappa + (n-2)\rho}{r + \kappa - \rho} b(\theta_k) - \frac{(n-1)\rho}{r + \kappa - \rho} \beta$$

so that \hat{b}_k is a θ_k -unbiased estimator of $x_{k,r+1}$ if and only if $n = 1$ or $\rho = 0$.

This verifies the assertion at the end of subsection 1K.

If $n = 1$ or $\rho = 0$, (2.14) reduces to (1.44). This is not unexpected. If $n = 1$, we obviously have the model of Example 1.2. $\rho = 0$ if and only if $C(c(\eta_1)) = 0$. Then we have $C(x_{kl}, x_{ij}) = 0$ for all $i \neq k$ and all j and l , that is, the claim amounts from different policies are uncorrelated. Then all the moments needed for our credibility estimator are the same as in Example 1.2, and consequently we get the same estimator.

We shall look at some asymptotic results.

When $r \rightarrow \infty$, we have $\frac{r}{r + \kappa - \rho} \rightarrow 1$, $\frac{r}{r + \kappa + (n-1)\rho} \rightarrow 1$,
 $\frac{\kappa + (n-1)\rho}{r + \kappa + (n-1)\rho} \rightarrow 0$, $\bar{x}_k \xrightarrow{a.s.} b(\theta_k)$, and $\bar{\bar{x}} \xrightarrow{a.s.} \frac{1}{n} \sum_{i=1}^n b(\theta_i)$.

From this follows that

$$\tilde{x}_{k,r+1} \xrightarrow{a.s.} b(\theta_k)$$

when $r \rightarrow \infty$, which is very satisfying.

If $\rho \neq 0$ and $n \rightarrow \infty$, we have $\frac{r}{r + \kappa + (n-1)\rho} \rightarrow 0$,

$\frac{\kappa + (n-1)\rho}{r + \kappa + (n-1)\rho} \rightarrow 1$, and $\bar{\bar{x}} \xrightarrow{a.s.} \frac{1}{r} \sum_{j=1}^r c(\eta_j) = \bar{c}$. This gives

$$(2.15) \quad \tilde{x}_{k,r+1} \xrightarrow{a.s.} \frac{r}{r + \kappa - \rho} (\bar{x}_k - \bar{c}) + \beta.$$

Rewriting (2.14) gives

$$(2.16) \quad \tilde{x}_{k,r+1} = \frac{r}{r + \kappa - \rho} \left[\bar{x}_k + \frac{n\rho}{r + \kappa + (n-1)\rho} (\beta - \bar{\bar{x}}) \right] + \frac{\kappa - \rho}{r + \kappa - \rho} \beta.$$

In (1.44) $\tilde{x}_{k,r+1}$ was a weighted sum of β and \bar{x}_k , the latter a reasonable estimator of $E(x_{k,r+1} | \theta_k)$. In (2.16) we can interpret

that \bar{x}_k has been replaced by $\bar{x}_k + \frac{n\rho}{r + \kappa + (n-1)\rho} (\beta - \bar{\bar{x}})$.

A correction term

$$h(\bar{\bar{x}}) = \frac{n\rho}{r + \kappa + (n-1)\rho} (\beta - \bar{\bar{x}})$$

has now entered to compensate for the collective random fluctuations caused by the η_j 's. We observe that $\frac{n\rho}{r + \kappa + (n-1)\rho}$

increases when ρ increases, that is, we need a greater correction when the variance of the $c(\eta_j)$'s increases, or less precisely, when

the collective seasonal fluctuations increase. We also have that $\frac{n \rho}{r + \kappa + (n - 1) \rho}$ decreases when r increases, that is, we need smaller correction when time increases.

Corresponding to (2.15) we have that

$$h(\bar{x}) \xrightarrow{a.s.} \bar{c}$$

when $n \rightarrow \infty$.

The above reasoning perhaps becomes clearer if we for a moment assume an additive model:

$$E(x_{ij} | \theta_i, \eta_j) = b(\theta_i) + d(\eta_j)$$

$$E(d(\eta_j)) = 0.$$

Then

$$E(h(\bar{x}) | \eta_1, \dots, \eta_r) = - \frac{n \rho}{r + \kappa + (n - 1) \rho} \frac{1}{r} \sum_{j=1}^r d(\eta_j)$$

and

$$h(\bar{x}) \xrightarrow{a.s.} - \frac{1}{r} \sum_{j=1}^r d(\eta_j)$$

when $n \rightarrow \infty$, and the nature of $h(\bar{x})$ as a mean to reduce the influence of η_j 's is clear.

From Theorem 1.2 we get the best linear unbiased homogeneous estimator of $x_{k,r+1}$

$$(2.17) \quad \ddot{x}_{k,r+1} = \frac{r}{r + \kappa - \rho} \bar{x}_k + \frac{\kappa - \rho}{r + \kappa - \rho} \bar{\bar{x}}.$$

2C. Inspired by the estimators of Bühlmann & Straub (1970) we shall develop reasonable estimators of κ and ρ . The idea is to

- i) find reasonable estimators of $EC(x_{11} | \theta_1)$, $C(c(\eta_1))$, and $C(b(\theta_1))$,

ii) replace $EC(x_{11}|\theta_1)$, $C(c(\eta_1))$, and $C(b(\theta_1))$ by their estimators in (2.10) and (2.11) to get estimators of κ and ρ ,

iii) replace κ and ρ in (2.17) by their estimators.

a) Estimating $EC(x_{11}|\theta_1)$.

$$e_1 = \frac{1}{n(r-1)} \sum_{i=1}^n \sum_{j=1}^r (x_{ij} - \bar{x}_i)^2$$

is an unbiased estimator of $EC(x_{11}|\theta_1)$.

b) Estimating $C(c(\eta_1))$.

$$(2.18) \quad e_2 = \frac{1}{n(n-1)(r-1)} \sum_{i=1}^n \sum_{\substack{k=1 \\ k \neq i}}^n \sum_{j=1}^r (x_{ij} - \bar{x}_i)(x_{kj} - \bar{x}_k)$$

is an unbiased estimator of $EC(x_{11}, x_{21}|\theta_1, \theta_2) = C(c(\eta_1))$.

Letting ${}_j\bar{x} = \frac{1}{n} \sum_{i=1}^n x_{ij}$, we have

$$e_2 = \frac{1}{n(n-1)(r-1)} \left[\sum_{i=1}^n \sum_{\substack{k=1 \\ k \neq i}}^n \sum_{j=1}^r (x_{ij} - \bar{x}_i)(x_{kj} - \bar{x}_k) - \sum_{i=1}^n \sum_{j=1}^r (x_{ij} - \bar{x}_i)^2 \right] =$$

$$\frac{1}{n-1} \left[\frac{1}{n(r-1)} \sum_{i=1}^n \sum_{k=1}^n \left(\sum_{j=1}^r x_{ij} x_{kj} - r \bar{x}_i \bar{x}_k \right) - e_1 \right] =$$

$$\frac{1}{n-1} \left[\frac{1}{n(r-1)} \left(\sum_{j=1}^r (n {}_j\bar{x})^2 - r (n \bar{\bar{x}})^2 \right) - e_1 \right] =$$

$$\frac{1}{n-1} \left[\frac{n}{r-1} \left(\sum_{j=1}^r {}_j\bar{x}^2 - r \bar{\bar{x}}^2 \right) - e_1 \right],$$

giving

giving

$$e_2 = \frac{1}{n-1} \left[\frac{n}{r-1} \sum_{j=1}^r (\bar{x}_j - \bar{\bar{x}})^2 - e_1 \right],$$

which is easier to compute than (2.18).

c) Estimating $C(b(\theta_1))$.

We have

$$\begin{aligned} E\left(\sum_{i=1}^n (\bar{x}_i - \bar{\bar{x}})^2\right) &= n E(\bar{x}_1 - \bar{\bar{x}})^2 = n C(\bar{x}_1 - \bar{\bar{x}}) = \\ n (C(\bar{x}_1) + C(\bar{\bar{x}}) - 2 C(\bar{x}_1, \bar{\bar{x}})) &= n (C(\bar{x}_1) - C(\bar{\bar{x}})) = \\ n (C(\bar{x}_1) - \frac{1}{n} C(\bar{x}_1) - (1 - \frac{1}{n}) C(\bar{x}_1, \bar{x}_2)) &= \\ (n-1) (C(\bar{x}_1) - C(\bar{x}_1, \bar{x}_2)) &= \\ (n-1) (C(b(\theta_1)) + \frac{1}{r} EC(x_{11}|\theta_1) - \frac{1}{r} C(c(n_1))) &= \\ (n-1) (C(b(\theta_1)) + \frac{1}{r} E(e_1 - e_2)). \end{aligned}$$

Hence

$$e_3 = \frac{1}{n-1} \left[\sum_{i=1}^n (\bar{x}_i - \bar{\bar{x}})^2 - \frac{1}{r} (e_1 - e_2) \right]$$

is an unbiased estimator of $C(b(\theta_1))$.

We now estimate κ with

$$\kappa^* = \frac{e_1}{e_3},$$

ρ with

$$\rho^* = \frac{e_2}{e_3},$$

and $\ddot{x}_{k,r+1}$ with

$$\ddot{x}_{k,r+1}^* = \frac{r}{r + \kappa^* - \rho^*} \bar{x}_k + \frac{\kappa^* - \rho^*}{r + \kappa^* - \rho^*} \bar{\bar{x}}.$$

Note that κ^* and ρ^* could be correlated with \bar{x}_k and \bar{x} so that $\ddot{x}_{k,r+1}^*$ is not generally unbiased.

2D. As a special case of our general model we briefly mentioned an additive model. Another special case that seems more natural in practice, is a multiplicative model where

$$E(x_{11} | \theta_1, \eta_1) = b(\theta_1) d(\eta_1)$$

$$E(d(\eta_1)) = 1.$$

Our linear credibility formulae seem natural in the additive model, not equally natural in the multiplicative case. And we could ask: Could we do better?

We shall now develop an alternative approach.

From subsection 1F we get that an estimator of $x_{k,r+1}$ is optimal if and only if it is an optimal estimator of $b(\theta_k)$.

What is now an optimal estimator of $b(\theta_k)$?

If we could observe the random variables $d(\eta_1), \dots, d(\eta_r)$,

it would be natural to base our estimator on $\frac{x_{k1}}{d(\eta_1)}, \dots, \frac{x_{kr}}{d(\eta_r)}$,

since $E\left(\frac{x_{kj}}{d(\eta_j)} \middle| \theta_k, \eta_j\right) = b(\theta_k)$. As $\frac{x_{k1}}{d(\eta_1)}, \dots, \frac{x_{kr}}{d(\eta_r)}$ are exchangeable relative to $b(\theta_k)$, the best linear inhomogeneous

estimator of $b(\theta_k)$ based on $\frac{x_{k1}}{d(\eta_1)}, \dots, \frac{x_{kr}}{d(\eta_r)}$ must be on the form

$$\alpha_0 + \alpha_1 \frac{1}{r} \sum_{j=1}^r \frac{x_{kj}}{d(\eta_j)},$$

where α_0 and α_1 are real constants.

Unfortunately we do not know $d(\eta_1), \dots, d(\eta_r)$. But they can be estimated. A natural estimator of $d(\eta_j)$ is the best linear

inhomogeneous estimator of $d(n_j)$ based on x_{1j}, \dots, x_{nj} , that is,

$$\tilde{d}_j = \frac{n}{n+v} \frac{j \bar{x}}{\beta} + \frac{v}{n+v}$$

with

$$v = \frac{EC(x_{11} | n_1)}{C(c(n_1))}.$$

Letting now $x_{kj}^* = \frac{x_{kj}}{\tilde{d}_j}$ for $j = 1, \dots, r$, and $\bar{x}_k^* = \frac{1}{r} \sum_{j=1}^r x_{kj}^*$, we

want the optimal estimator of $x_{k,r+1}$ of the form $g_0 + g_1 \bar{x}_k^*$. From (1.6) and (1.7) follows that this optimal estimator is

$$\tilde{x}_{k,r+1} = \gamma_0 + \gamma_1 \bar{x}_k^*$$

with

$$\gamma_0 = E(x_{11}) - \gamma_1 E(x_{11}^*)$$

and

$$\gamma_1 = \frac{C(\bar{x}_k^*, x_{k,r+1})}{C(\bar{x}_k^*)}.$$

As

$$C(\bar{x}_k^*, x_{k,r+1}) = C(x_{11}^*, x_{12})$$

and

$$C(\bar{x}_k^*) = \frac{1}{r} C(x_{11}^*) + (1 - \frac{1}{r}) C(x_{11}^*, x_{12}^*)$$

these quantities can easily be estimated.

For closer examination of the behaviour of $\tilde{x}_{k,r+1}$ simulation is suggested.

2E. The multiplicative model has also been treated by Welten (1968). He uses the premium (1.44). If the claim amounts in j -th term are small, this could be due to small $d(n_j)$. This would lead to too small premiums to cover the expected future claims. However, the small claim amounts in j -th

term have given the insurance company a profit, and Welten argues that a part of this profit should be paid into a "bonus reserve" to cover the future claims.

2F. Both $\tilde{x}_{k,r+1}$, $\ddot{x}_{k,r+1}$, and $\tilde{\tilde{x}}_{k,r+1}$ (and Welten's approach) are developed under the very strict assumption that our portfolio consists of n policies that have been running in the same r terms. It is very unlikely that we shall meet this situation in reality, and we therefore need a less restrictive model.

Unfortunately, in subsection 2B we would lose some of the exchangeability properties and get a messy system of linear equations, solvable, but too complicated to be of any practical use. The present author believes that instead of developing the best linear inhomogeneous and unbiased homogeneous formulae, one ought to examine the structure of (2,14) and (2,17) and try to develop similar, not too complicated formulae under the more general conditions.

It seems that such generalizations would be somewhat easier in the approach of subsection 2E.

3. A hiererchical credibility regression model.

A regression model with random parameters on two levels is developed, from which a model by Taylor (1974) and Jewell (1975c,d) is derived as a special case.

3A. In Example 1.2 assume that all the policies are taken from a certain district. If we compare the claim amounts from this district with claim amounts from another district, there may be systematic differences. We shall explain these by assuming that each district is characterized by an unknown random parameter. Given this random parameter the policies of the district are independent.

Example 1.2 was derived as a special case of the model of subsection 1I. We shall generalize the model of subsection 1I according to the above remarks. The insurance model will be treated as a special case (Example 3.1).

3B. Let \underline{x}_1 and \underline{x}_2 be observable random vectors of dimensions $r_1 \times 1$ resp. $r_2 \times 1$. We want to estimate an unknown random variable m . Let θ and η be two unknown random parameters. We assume that \underline{x}_2 and θ are independent given η , and that m is independent of \underline{x}_1 and \underline{x}_2 given θ and η . We further assume that

$$E(\underline{x}_1 | \theta, \eta) = \underline{Y}_1 \underline{b}(\theta, \eta)$$

$$E(m | \theta, \eta) = \underline{a}' \underline{b}(\theta, \eta)$$

$$E(\underline{b}(\theta, \eta) | \eta) = \underline{\beta}(\eta)$$

$$E(\underline{x}_2 | \eta) = \underline{Y}_2 \underline{\beta}(\eta)$$

Here \underline{Y}_i is a non-random $r_i \times q$ -matrix for $i = 1, 2$, \underline{a} a non-random $q \times 1$ -vector, and $\underline{b}(\theta, \eta)$ and $\underline{\beta}(\eta)$ $q \times 1$ vector functions of the

random parameters. \underline{Y}_1 has rank q .

Let

$$\underline{\beta} = E(\underline{\beta}(\eta))$$

$$\underline{\Lambda} = EC(\underline{b}(\theta, \eta) | \eta)$$

$$\underline{Y} = (\underline{Y}_1', \underline{Y}_2')'$$

$$\underline{x} = (\underline{x}_1', \underline{x}_2')'$$

We assume that

$$\underline{\Xi} = C(\underline{\beta}(\eta))$$

and

$$\underline{\Sigma} = EC(\underline{x} | \eta)$$

are invertible.

Let

$$(3.1) \quad \tilde{m} = \gamma_0 + \underline{Y}' \underline{x}$$

be the best linear inhomogeneous estimator of m based on \underline{x} .

(1.4) and (1.5) give

$$(3.2) \quad \gamma_0 = E(m) - \underline{Y}' E(\underline{x})$$

$$(3.3) \quad \underline{Y}' C(\underline{x}) = C(m, \underline{x}')$$

We have

$$(3.4) \quad C(\underline{x}) = \underline{\Sigma} + \underline{Y} \underline{\Xi} \underline{Y}'$$

$$(3.5) \quad C(m, \underline{x}') = \underline{a}' [\underline{\Lambda} (\underline{Y}_1', 0) + \underline{\Xi} \underline{Y}']$$

By putting (3.4) and (3.5) into (3.3) we get

$$(3.6) \quad \underline{Y}' (\underline{\Sigma} + \underline{Y} \underline{\Xi} \underline{Y}') = \underline{a}' [\underline{\Lambda} (\underline{Y}_1', 0) + \underline{\Xi} \underline{Y}']$$

Multiplying (3.6) by $\underline{\Sigma}^{-1} \underline{Y}$ gives

$$(3.7) \quad \underline{Y}' \underline{Y} (\underline{I} + \underline{\Xi} \underline{Y}' \underline{\Sigma}^{-1} \underline{Y}) = \underline{a}' (\underline{Z} + \underline{\Xi} \underline{Y}' \underline{\Sigma}^{-1} \underline{Y})$$

with

$$(3.8) \quad \underline{Z} = \underline{\Lambda} (\underline{Y}'_1, \underline{0}) \underline{\Sigma}^{-1} \underline{Y}.$$

Zellner (1971, p.231) states that if \underline{B} is a $k \times l$ -matrix and \underline{C} an $l \times k$ -matrix, then $|\underline{I} + \underline{B} \underline{C}| = |\underline{I} + \underline{C} \underline{B}|$. This gives

$$|\underline{I} + \underline{\Xi} \underline{Y}' \underline{\Sigma}^{-1} \underline{Y}| = |\underline{I} + \underline{Y} \underline{\Xi} \underline{Y}' \underline{\Sigma}^{-1}| = |(\underline{\Sigma} + \underline{Y} \underline{\Xi} \underline{Y}') \underline{\Sigma}^{-1}| = |\underline{\Sigma} + \underline{Y} \underline{\Xi} \underline{Y}'| |\underline{\Sigma}^{-1}|.$$

As $\underline{\Sigma}$ is positive definite and $\underline{Y} \underline{\Xi} \underline{Y}'$ non-negative definite, $\underline{\Sigma} + \underline{Y} \underline{\Xi} \underline{Y}'$ must be positive definite and has thereby non-zero determinant. Accordingly $\underline{I} + \underline{\Xi} \underline{Y}' \underline{\Sigma}^{-1} \underline{Y}$ has non-zero determinant and is thereby invertible.

Hence we get from (3.7)

$$(3.9) \quad \underline{Y}' \underline{Y} = \underline{a}' (\underline{Z} + \underline{\Xi} \underline{Y}' \underline{\Sigma}^{-1} \underline{Y}) (\underline{I} + \underline{\Xi} \underline{Y}' \underline{\Sigma}^{-1} \underline{Y})^{-1}.$$

From (3.6) we get

$$\underline{Y}' = \{ \underline{a}' [\underline{\Lambda} (\underline{Y}'_1, \underline{0}) + \underline{\Xi} \underline{Y}'] - \underline{Y}' \underline{Y} \underline{\Xi} \underline{Y}' \} \underline{\Sigma}^{-1},$$

giving

$$(3.10) \quad \underline{Y}' = \underline{a}' \underline{\Lambda} (\underline{Y}'_1, \underline{0}) \underline{\Sigma}^{-1} + (\underline{a}' - \underline{Y}' \underline{Y}) \underline{\Xi} \underline{Y}' \underline{\Sigma}^{-1}.$$

By multiplying (3.10) by \underline{x} and using (3.8) we get

$$(3.11) \quad \underline{Y}' \underline{x} = \underline{a}' \underline{Z} \hat{\underline{b}} + (\underline{a}' - \underline{Y}' \underline{Y}) \underline{\Xi} \underline{Y}' \underline{\Sigma}^{-1} \underline{Y} \hat{\underline{\beta}}$$

with

$$\hat{\underline{b}} = [(\underline{Y}'_1, \underline{0}) \underline{\Sigma}^{-1} \underline{Y}]^{-1} (\underline{Y}'_1, \underline{0}) \underline{\Sigma}^{-1} \underline{x}$$

and

$$(3.12) \quad \hat{\underline{\beta}} = (\underline{Y}' \underline{\Sigma}^{-1} \underline{Y})^{-1} \underline{Y}' \underline{\Sigma}^{-1} \underline{x}.$$

Putting (3.9) into (3.11) gives

$$\begin{aligned} \underline{Y}' \underline{x} &= \underline{a}' \{ \underline{Z} \hat{\underline{b}} + [\underline{I} - (\underline{Z} + \underline{\Xi} \underline{Y}' \underline{\Sigma}^{-1} \underline{Y}) (\underline{I} + \underline{\Xi} \underline{Y}' \underline{\Sigma}^{-1} \underline{Y})^{-1}] \underline{\Xi} \underline{Y}' \underline{\Sigma}^{-1} \underline{Y} \hat{\underline{\beta}} \} = \\ &= \underline{a}' [\underline{Z} \hat{\underline{b}} + (\underline{I} - \underline{Z}) (\underline{I} + \underline{\Xi} \underline{Y}' \underline{\Sigma}^{-1} \underline{Y})^{-1} \underline{\Xi} \underline{Y}' \underline{\Sigma}^{-1} \underline{Y} \hat{\underline{\beta}}], \end{aligned}$$

and we get

$$(3.13) \quad \underline{y}'\underline{x} = \underline{a}'[\underline{Z} \hat{\underline{b}} + (\underline{I} - \underline{Z}) \underline{\Delta} \hat{\underline{\beta}}]$$

with

$$(3.14) \quad \underline{\Delta} = (\underline{I} + \underline{E} \underline{Y}' \underline{\Sigma}^{-1} \underline{Y})^{-1} \underline{E} \underline{Y}' \underline{\Sigma}^{-1} \underline{Y}.$$

(3.1), (3.2), and (3.13) now finally give

$$(3.15) \quad \tilde{\underline{m}} = \underline{a}'\{\underline{Z} \hat{\underline{b}} + (\underline{I} - \underline{Z}) [\underline{\Delta} \hat{\underline{\beta}} + (\underline{I} - \underline{\Delta}) \underline{\beta}]\}.$$

(The above derivations are very similar to derivations by Hachemeister (1975) and Taylor (1977).)

Let \underline{m} be an unknown random $s \times 1$ -vector, independent of \underline{x} given θ and η , and with

$$E(\underline{m}|\theta, \eta) = \underline{A} \underline{b}(\theta, \eta),$$

where \underline{A} is a non-random $s \times q$ -matrix. Then the vector generalization

$$(3.16) \quad \tilde{\underline{m}} = \underline{A}\{\underline{Z} \hat{\underline{b}} + (\underline{I} - \underline{Z}) [\underline{\Delta} \hat{\underline{\beta}} + (\underline{I} - \underline{\Delta}) \underline{\beta}]\}$$

of (3.15) is obvious.

Remarks.

i) (3.14) gives

$$\underline{\Delta} = \underline{I} - (\underline{I} + \underline{E} \underline{Y}' \underline{\Sigma}^{-1} \underline{Y})^{-1} = \underline{E} \underline{Y}' \underline{\Sigma}^{-1} \underline{Y} (\underline{I} + \underline{E} \underline{Y}' \underline{\Sigma}^{-1} \underline{Y})^{-1}.$$

By comparing with (1.22) and (1.30) we see that

$$\tilde{\underline{\beta}} = \underline{\Delta} \hat{\underline{\beta}} + (\underline{I} - \underline{\Delta}) \underline{\beta}$$

is the best linear inhomogeneous estimator of $\underline{\beta}(\eta)$ based on \underline{x} .

We can now write

$$(3.17) \quad \tilde{\underline{m}} = \underline{A} [\underline{Z} \hat{\underline{b}} + (\underline{I} - \underline{Z}) \tilde{\underline{\beta}}],$$

ii) Suppose that $\underline{\beta}(\eta) = \underline{\beta}$. Then $\underline{E} = \underline{0}$, and we get

$$\tilde{\underline{m}} = \underline{A} [\underline{Z} \hat{\underline{b}} + (\underline{I} - \underline{Z}) \underline{\beta}],$$

This result is intuitively very sound, $\hat{\underline{\beta}}$ is the best linear η - unbiased estimator of $\underline{\beta}(\eta)$. But now we know $\underline{\beta}(\eta)$, and consequently $\hat{\underline{\beta}}$ drops out.

Under the stronger assumption that η is a constant, our model reduces to the model of subsection 1I.

iii) Suppose now on the other hand that our prior knowledge about the distribution of $\underline{\beta}(\eta)$ is extremely vague. This could be formalized by putting the precision matrix of $\underline{\beta}(\eta)$, $\underline{\Xi}^{-1}$, equal to $\underline{0}$.

From (3.14) we get

$$(3.18) \quad \underline{\Delta} = (\underline{\Xi}^{-1} + \underline{Y}'\underline{\Sigma}^{-1}\underline{Y})^{-1}\underline{Y}'\underline{\Sigma}^{-1}\underline{Y}.$$

When $\underline{\Xi}^{-1} = \underline{0}$, (3.18) gives $\underline{\Delta} = \underline{I}$, and we get $\underline{\tilde{m}} = \underline{A} [\underline{Z} \hat{\underline{b}} + (\underline{I} - \underline{Z}) \hat{\underline{\beta}}]$.

Our knowledge of the distribution of $\underline{\beta}(\eta)$ is now so vague that we are not willing to put any weight on its expectation $\underline{\beta}$.

3C. As $\hat{\underline{\beta}}$ is the best linear unbiased estimator of $\underline{\beta}$, we get the best linear unbiased homogeneous estimator of \underline{m}

$$\underline{\tilde{m}} = \underline{A} [\underline{Z} \hat{\underline{b}} + (\underline{I} - \underline{Z}) \hat{\underline{\beta}}]$$

from Theorem 1.2 and (3.16),

Suppose now that we have an additional sample \underline{x}_3 independent of \underline{x} , \underline{m} , θ , and η . We assume that

$$E(\underline{x}_3) = \underline{Y}_3 \underline{\beta},$$

where \underline{Y}_3 is a non-random matrix.

Then the best linear unbiased homogeneous estimator of \underline{m} based on \underline{x} is

$$\underline{\tilde{m}} = \underline{A} \{ \underline{Z} \hat{\underline{b}} + (\underline{I} - \underline{Z}) [\underline{\Delta} \hat{\underline{\beta}} + (\underline{I} - \underline{\Delta}) \hat{\underline{\beta}}] \},$$

where $\hat{\underline{\beta}}$ is the best linear unbiased estimator of $\underline{\beta}$ based on \underline{x} and \underline{x}_3 .

Corresponding to remark i) of subsection 3B we see that

$$\ddot{\underline{\beta}} = \underline{A} \hat{\underline{\beta}} + (\underline{I} - \underline{A}) \hat{\underline{\beta}}$$

is the best linear unbiased homogeneous estimator of $\underline{\beta}(\eta)$ and get

$$\ddot{\underline{m}} = \underline{A} [\underline{Z} \hat{\underline{b}} + (\underline{I} - \underline{Z}) \ddot{\underline{\beta}}]$$

corresponding to (3.17).

3D. We now make the assumption that \underline{x}_1 and \underline{x}_2 are independent given η . We then get

$$\underline{\Sigma} = \begin{pmatrix} \underline{\Sigma}_1 & \underline{0} \\ \underline{0} & \underline{\Sigma}_2 \end{pmatrix},$$

where

$$\underline{\Sigma}_i = EC(\underline{x}_i | \eta)$$

for $i = 1, 2$.

As in subsection 1J we get the simplifications

$$(3.19) \quad \hat{\underline{b}} = (\underline{Y}_1' \underline{\Sigma}_1^{-1} \underline{Y}_1)^{-1} \underline{Y}_1' \underline{\Sigma}_1^{-1} \underline{x}_1$$

and

$$(3.20) \quad \underline{Z} = \underline{A} \underline{Y}_1' \underline{\Sigma}_1^{-1} \underline{Y}_1$$

and observe that $\hat{\underline{b}}$ is the best linear (θ, η) -unbiased estimator of $\underline{b}(\theta, \eta)$.

Corresponding to the remark in subsection 1K, we observe that the optimal linear estimator of $\underline{b}(\theta, \eta)$ is a weighted average of the best linear (θ, η) -unbiased estimator of $\underline{b}(\theta, \eta)$ and the best linear inhomogeneous estimator of $E(\underline{b}(\theta, \eta) | \eta)$ (inhomogeneous case), or the best linear (θ, η) -unbiased estimator of $\underline{b}(\theta, \eta)$ and the best linear unbiased homogeneous estimator of $E(\underline{b}(\theta, \eta) | \eta)$ (unbiased homogeneous case). The weights are the same in both cases.

The optimal estimator of \underline{m} is obtained by multiplying \underline{A} by the optimal estimator of $\underline{b}(\theta, \eta)$.

Example 3.1. Assume that we have an insurance portfolio consisting of n policies. x_{ij} is the total claim amount of i -th policy in the j -th term it is running. Our portfolio is characterized by an unknown random parameter η . Given this parameter, claim data from different policies are independent. Furthermore each policy i is characterized by an unknown random parameter θ_i . We assume that $\theta_1, \dots, \theta_n$ are conditionally independent and identically distributed given η , and that x_{i1}, x_{i2}, \dots are independent and identically distributed given θ_i and η with common conditional cumulative distribution on the form $F(\cdot | \theta_i, \eta)$ with F independent of i . i -th policy has been running in r_i terms.

Let

$$b(\theta_1, \eta) = E(x_{11} | \theta_1, \eta)$$

$$\beta(\eta) = E(b(\theta_1, \eta) | \eta)$$

$$\beta = E(\beta(\eta)).$$

$$\varphi = EC(x_{11} | \theta_1, \eta).$$

It is assumed that

$$\lambda = EC(b(\theta_1, \eta) | \eta)$$

and

$$\xi = C(\beta(\eta))$$

are non-zero.

We want to estimate x_{k, r_k+1} by \tilde{x}_{k, r_k+1} , the best linear inhomogeneous estimator based on the observed claim amounts from the portfolio.

From (3.15) and (3.19) we get that \tilde{x}_{k,r_k+1} can be written on the form

$$(3.21) \quad \tilde{x}_{k,r_k+1} = \zeta_k \hat{b}_k + (1 - \zeta_k) [\delta \hat{\beta} + (1 - \delta) \beta].$$

Here ζ_k and δ are constants. \hat{b}_k is the best linear (θ_k, η) -unbiased estimator of $b(\theta_k, \eta)$, $\hat{\beta}$ the best linear η -unbiased estimator of $\beta(\eta)$, and

$$\tilde{\beta} = \delta \hat{\beta} + (1 - \delta) \beta$$

the best linear inhomogeneous estimator of $\beta(\eta)$.

It is obvious that

$$(3.22) \quad \hat{b}_k = \bar{x}_k = \frac{1}{r_k} \sum_{j=1}^{r_k} x_{kj}.$$

By comparing (3.20) with (1.37) and the present model with the model of Example 1.2 we get

$$(3.23) \quad \zeta_k = \frac{n_k}{n_k + \kappa}$$

with $\kappa = \frac{\varphi}{\lambda}$.

Similarly by comparing (3.12) with (1.11) and the present model with the model of Example 1.2 we get

$$(3.24) \quad \hat{\beta} = \frac{\sum_{i=1}^n \frac{r_i}{r_i + \kappa} \bar{x}_i}{\sum_{i=1}^n \frac{r_i}{r_i + \kappa}} = \sum_{i=1}^n v_i \bar{x}_i$$

with

$$v_i = \frac{\frac{r_i}{r_i + \kappa}}{\sum_{l=1}^n \frac{r_l}{r_l + \kappa}}.$$

It remains to determine δ . But since $\tilde{\beta}$ is the best linear inhomogeneous estimator of $\beta(\eta)$, (1.6) gives

$$\delta = \frac{C(\beta(\eta), \hat{\beta})}{C(\hat{\beta})} = \frac{EC(\beta(\eta))}{EC(\hat{\beta}|\eta) + CE(\hat{\beta}|\eta)} = \frac{\xi}{\sum_{i=1}^n v_i^2 EC(\bar{x}_i|\eta) + C(\beta(\eta))} =$$

$$\frac{\xi}{\sum_{i=1}^n v_i^2 [EC(\bar{x}_i|\theta_i, \eta) + EC(E(\bar{x}_i|\theta_i, \eta)|\eta)] + \xi} =$$

$$\frac{\xi}{\sum_{i=1}^n v_i^2 \left[\frac{1}{r_i} EC(x_{i1}|\theta_i, \eta) + EC(b(\theta_i, \eta)|\eta) \right] + \xi} =$$

$$\frac{\xi}{\sum_{i=1}^n v_i^2 \left(\frac{\varphi}{r_i} + \lambda \right) + \xi} = \frac{\xi}{\sum_{i=1}^n v_i^2 \lambda \left(\frac{\kappa}{r_i} + 1 \right) + \xi} =$$

$$\frac{\xi}{\lambda \sum_{i=1}^n v_i^2 \frac{r_i + \kappa}{r_i} + \xi} = \frac{\xi}{\frac{\lambda}{\sum_{i=1}^n \frac{r_i}{r_i + \kappa}} + \xi},$$

giving

$$(3.25) \quad \delta = \frac{\sum_{i=1}^n \frac{r_i}{r_i + \kappa}}{\sum_{i=1}^n \frac{r_i}{r_i + \kappa} + \rho}$$

with $\rho = \frac{\lambda}{\xi}$.

From (3.21), (3.22), (3.23), (3.24), and (3.25) we now get

$$(3.26) \quad \tilde{x}_{k,r_k+1} = \frac{r_k}{r_k + \kappa} \bar{x}_k + \frac{\kappa}{r_k + \kappa} \left(\frac{\sum_{i=1}^n \frac{r_i}{r_i + \kappa}}{\sum_{i=1}^n \frac{r_i}{r_i + \kappa} + \rho} \frac{\sum_{i=1}^n \frac{\kappa}{r_i + \kappa} \bar{x}_i}{\sum_{i=1}^n \frac{r_i}{r_i + \kappa}} + \frac{\frac{\rho}{\sum_{i=1}^n \frac{r_i}{r_i + \kappa} + \rho} \beta \right).$$

When $r_1 = r_2 = \dots = r_n = r$, (3.26) reduces to

$$\tilde{x}_{k,r+1} = \frac{r}{r + \kappa} \bar{x}_k + \frac{\kappa}{r + \kappa} \frac{\frac{n}{r + \kappa} \bar{x}}{n \frac{r}{r + \kappa} + \rho} + \frac{\rho}{n \frac{r}{r + \kappa} + \rho} \beta$$

with $\bar{x} = \frac{1}{n} \sum_{i=1}^n \bar{x}_i$, or

$$\tilde{x}_{k,r+1} = \frac{r}{r + \kappa} \bar{x}_k + \frac{\kappa}{r + \kappa} \left(\frac{n}{n + \pi(r)} \bar{x} + \frac{\pi(r)}{n + \pi(r)} \beta \right)$$

with $\pi(r) = \rho \left(1 + \frac{\kappa}{r} \right)$.

The model of Example 3.1 has previously been studied by Taylor (1974) and Jewell (1975c,d).

□

4. Optimal choice of observators.

The question on what observators to base our credibility estimators, is discussed in a general multi-dimensional model. Concepts of sufficiency, completeness, θ -sufficiency, and θ -completeness are often useful in our search for good observators. Some of the present results are closely related to results by Taylor (1977).

4A. Assume that we want to estimate an unknown random variable m by an observable random $q \times 1$ -vector \underline{x} . We assume that \underline{x} and m are independent given an unknown random parameter θ , and that

$$E(m|\theta) = \underline{a}'\underline{b}(\theta),$$

where \underline{a} is a non-random $q \times 1$ -vector and $\underline{b}(\theta)$ a $q \times 1$ vector function of θ . We assume that $C(\underline{b}(\theta))$ is invertible.

Under certain conditions (cfr. subsection 1K), which we assume hold, the best linear inhomogeneous estimator of m based on \underline{x} can be written

$$\tilde{m} = \underline{a}'[\underline{Z} \hat{\underline{b}} + (\underline{I} - \underline{Z}) \underline{\beta}]$$

and the best linear unbiased homogeneous estimator

$$\ddot{m} = \underline{a}'[\underline{Z} \hat{\underline{b}} + (\underline{I} - \underline{Z}) \hat{\underline{\beta}}].$$

Here \underline{Z} is a non-random $q \times q$ -matrix, $\underline{\beta} = E(\underline{b}(\theta))$, $\hat{\underline{b}}$ the best linear θ -unbiased estimator of $\underline{b}(\theta)$, and $\hat{\underline{\beta}}$ the best linear unbiased estimator of $\underline{\beta}$.

As the non-random vector $\underline{\beta}$ obviously has expectation $\underline{\beta}$, we shall allow $\hat{\underline{\beta}}$ to be called an unbiased estimator of $\underline{\beta}$. This convention enables us to develop results for inhomogeneous and unbiased homogeneous estimators at the same time.

We now have that both \tilde{m} and \ddot{m} can be written on the form

$$\underline{g}'\underline{b}^* + (\underline{a} - \underline{g})'\underline{\beta}^*,$$

where \underline{b}^* is a θ -unbiased estimator of $\underline{b}(\theta)$, $\underline{\beta}^*$ an unbiased estimator of $\underline{\beta}$ and \underline{g} a non-random $q \times 1$ -vector.

As an estimator of m is optimal if and only if it is an optimal estimator of $E(m|\theta) = \underline{a}'\underline{b}(\theta)$, we shall in the sequel assume that $m = \underline{a}'\underline{b}(\theta)$.

4B. Now let \underline{b}^* be a θ -unbiased estimator of $\underline{b}(\theta)$ and $\underline{\beta}^*$ an unbiased estimator of $\underline{\beta}$. We want to determine the best estimator of m of the form

$$(4.1) \quad \underline{g}'\underline{b}^* + (\underline{a} - \underline{g})'\underline{\beta}^*.$$

By writing

$$E(\underline{g}'\underline{b}^* + (\underline{a} - \underline{g})'\underline{\beta}^* - m)^2 = E(\underline{g}'(\underline{b}^* - \underline{\beta}^*) - \underline{a}'[\underline{b}(\theta) - \underline{\beta}^*])^2$$

we observe that $\underline{y}'\underline{b}^* + (\underline{a} - \underline{y})'\underline{\beta}^*$ is the best estimator of m of the form (4.1) if and only if $\underline{y}'(\underline{b}^* - \underline{\beta}^*)$ is the best linear inhomogeneous estimator of $\underline{a}'[\underline{b}(\theta) - \underline{\beta}^*]$

based on $\underline{b}^* - \underline{\beta}^*$. If we assume that $C(\underline{b}^* - \underline{\beta}^*)$ is invertible,

(1.6) gives

$$\underline{y}' = C(\underline{a}'[\underline{b}(\theta) - \underline{\beta}^*], (\underline{b}^* - \underline{\beta}^*)') C(\underline{b}^* - \underline{\beta}^*)^{-1} = \underline{a}'\underline{Z}(\underline{b}^*, \underline{\beta}^*)$$

with

$$(4.2) \quad \underline{Z}(\underline{b}^*, \underline{\beta}^*) = C(\underline{b}(\theta) - \underline{\beta}^*, (\underline{b}^* - \underline{\beta}^*)') C(\underline{b}^* - \underline{\beta}^*)^{-1}.$$

This gives that the optimal estimator of m of the form (4.1) is

$$\ddot{m}(\underline{b}^*, \underline{\beta}^*) = \underline{a}'\{\underline{Z}(\underline{b}^*, \underline{\beta}^*) \underline{b}^* + [\underline{I} - \underline{Z}(\underline{b}^*, \underline{\beta}^*)] \underline{\beta}^*\}.$$

From (4.2) we see that $\underline{Z}(\underline{b}(\theta), \underline{\beta}^*) = \underline{I}$, giving

$\ddot{m}(\underline{b}(\theta), \underline{\beta}^*) = \underline{a}'\underline{b}(\theta) = m$. This result is obvious; m is definitely the best estimator of m .

In the inhomogeneous case with $\underline{\beta}^* = \underline{\beta}$ we have

$$C(\underline{b}^* - \underline{\beta}) = C(\underline{b}^*) = C(\underline{b}(\theta)) + EC(\underline{b}^*|\theta).$$

As $C(\underline{b}(\theta))$ was assumed invertible and thereby positive definite, and $EC(\underline{b}^*|\theta)$ is non-negative definite, $C(\underline{b}^*)$ is invertible whenever it exists. We get

$$\underline{Z}(\underline{b}^*, \underline{\beta}) = C(\underline{b}(\theta)) [C(\underline{b}(\theta)) + EC(\underline{b}^*|\theta)]^{-1}.$$

4C. What \underline{b}^* and $\underline{\beta}^*$ are optimal to use? To answer this question we have to examine the mean square $E(\hat{m}(\underline{b}^*, \underline{\beta}^*) - m)^2$.

We have

$$E(\hat{m}(\underline{b}^*, \underline{\beta}^*) - m)^2 =$$

$$C(\underline{a}' \{ \underline{Z}(\underline{b}^*, \underline{\beta}^*) \underline{b}^* + [\underline{I} - \underline{Z}(\underline{b}^*, \underline{\beta}^*)] \underline{\beta}^* \} - \underline{a}' \underline{b}(\theta)) =$$

$$\underline{a}' C(\underline{Z}(\underline{b}^*, \underline{\beta}^*) (\underline{b}^* - \underline{\beta}^*) - (\underline{b}(\theta) - \underline{\beta}^*)) \underline{a},$$

giving

$$E(\hat{m}(\underline{b}^*, \underline{\beta}^*) - m)^2 = \underline{a}' \underline{\Pi}(\underline{b}^*, \underline{\beta}^*) \underline{a}$$

with

$$\underline{\Pi}(\underline{b}^*, \underline{\beta}^*) = C(\underline{Z}(\underline{b}^*, \underline{\beta}^*) (\underline{b}^* - \underline{\beta}^*) - (\underline{b}(\theta) - \underline{\beta}^*)).$$

The matrix function $\underline{\Pi}(\cdot, \cdot)$ (not really a function of estimators, but of distributions of estimators) seems to play an important part in choice of observators \underline{b}^* and $\underline{\beta}^*$. We observe that if \underline{b}^{**} is a θ -unbiased estimator of $\underline{b}(\theta)$ and $\underline{\beta}^{**}$ an unbiased estimator of $\underline{\beta}$, and $\underline{\Pi}(\underline{b}^*, \underline{\beta}^*)$ is less than $\underline{\Pi}(\underline{b}^{**}, \underline{\beta}^{**})$, then

$$\underline{c}' \{ \underline{Z}(\underline{b}^*, \underline{\beta}^*) \underline{b}^* + [\underline{I} - \underline{Z}(\underline{b}^*, \underline{\beta}^*)] \underline{\beta}^* \}$$

is a not worse estimator of $\underline{c}' \underline{b}(\theta)$ than

$$\underline{c}' \{ \underline{Z}(\underline{b}^{**}, \underline{\beta}^{**}) \underline{b}^{**} + [\underline{I} - \underline{Z}(\underline{b}^{**}, \underline{\beta}^{**})] \underline{\beta}^{**} \}$$

for all non-random vectors $\underline{c} \in \mathbb{R}^q$ and better for some \underline{c} .

After these remarks we are interested in finding sufficient conditions for $\underline{\Pi}(\underline{b}^*, \underline{\beta}^*)$ to be less than $\underline{\Pi}(\underline{b}^{**}, \underline{\beta}^{**})$. We further

want to find when there exist a \underline{b}^* and $\underline{\beta}^*$ minimizing $\Pi(\cdot, \cdot)$.

We have

$$\begin{aligned} \Pi(\underline{b}^*, \underline{\beta}^*) &= C(\underline{b}(\theta) - \underline{\beta}^*) + Z(\underline{b}^*, \underline{\beta}^*) C(\underline{b}^* - \underline{\beta}^*) Z(\underline{b}^*, \underline{\beta}^*)' - \\ &C(\underline{b}(\theta) - \underline{\beta}^*, (\underline{b}^* - \underline{\beta}^*)') Z(\underline{b}^*, \underline{\beta}^*)' - Z(\underline{b}^*, \underline{\beta}^*) C(\underline{b}^* - \underline{\beta}^*, (\underline{b}(\theta) - \underline{\beta}^*)'). \end{aligned}$$

By using (4.2) we get

$$\begin{aligned} (4.3) \quad \Pi(\underline{b}^*, \underline{\beta}^*) &= C(\underline{b}(\theta) - \underline{\beta}^*) - \\ &C(\underline{b}(\theta) - \underline{\beta}^*, (\underline{b}^* - \underline{\beta}^*)') C(\underline{b}^* - \underline{\beta}^*)^{-1} C(\underline{b}(\theta) - \underline{\beta}^*, (\underline{b}(\theta) - \underline{\beta}^*)'). \end{aligned}$$

As this expression seems a bit complicated to discuss in general, we shall in the next subsection put some restrictions on possible observators \underline{b}^* and $\underline{\beta}^*$.

4D. Let B_1 be a set of θ -unbiased estimators of $\underline{b}(\theta)$ and B_2 a set of unbiased estimators of $\underline{\beta}$. We assume that B_2 is stochastically independent of B_1 and θ , that $C(\underline{b}^*)$ exists for all $\underline{b}^* \in B_1$, and that $C(\underline{\beta}^*)$ exists for all $\underline{\beta}^* \in B_2$.

For $\underline{b}^* \in B_1$ and $\underline{\beta}^* \in B_2$ we have

$$(4.4) \quad C(\underline{b}^* - \underline{\beta}^*) = C(\underline{b}^*) + C(\underline{\beta}^*)$$

$$(4.5) \quad C(\underline{b}(\theta) - \underline{\beta}^*) = C(\underline{b}(\theta) - \underline{\beta}^*, (\underline{b}^* - \underline{\beta}^*)') = C(\underline{b}(\theta)) + C(\underline{\beta}^*)$$

$$(4.6) \quad C(\underline{b}^*) = C(\underline{b}(\theta)) + EC(\underline{b}^* | \theta).$$

From (4.4) and (4.6) follows easily that $C(\underline{b}^* - \underline{\beta}^*)$ is invertible.

(4.3), (4.4), (4.5), and (4.6) give

$$\begin{aligned} (4.7a) \quad \Pi(\underline{b}^*, \underline{\beta}^*) &= C(\underline{b}(\theta)) + C(\underline{\beta}^*) - \\ &[C(\underline{b}(\theta)) + C(\underline{\beta}^*)][C(\underline{b}^*) + C(\underline{\beta}^*)]^{-1} [C(\underline{b}(\theta)) + C(\underline{\beta}^*)] = \end{aligned}$$

$$(4.7b) \quad [C(\underline{b}(\theta)) + C(\underline{\beta}^*)][C(\underline{b}^*) + C(\underline{\beta}^*)]^{-1} EC(\underline{b}^* | \theta) =$$

$$(4.7c) \quad EC(\underline{b}^* | \theta) - EC(\underline{b}^* | \theta) [C(\underline{b}^*) + C(\underline{\beta}^*)]^{-1} EC(\underline{b}^* | \theta).$$

We are now able to prove two theorems.

Theorem 4.1. Let \underline{b}^* and $\underline{b}^{**} \in B_1$ and $\underline{g}^* \in B_2$. Then

$\Pi(\underline{b}^*, \underline{g}^*)$ is less than $\Pi(\underline{b}^{**}, \underline{g}^*)$ if and only if $C(\underline{b}^*)$ is less than $C(\underline{b}^{**})$.

Proof. From (4.7a) follows that $\Pi(\underline{b}^*, \underline{g})$ is less than $\Pi(\underline{b}^{**}, \underline{g})$ if and only if

$$[C(\underline{b}(\theta)) + C(\underline{g}^*)][C(\underline{b}^{**}) + C(\underline{g}^*)]^{-1} [C(\underline{b}(\theta)) + C(\underline{g}^*)]$$

is less than

$$[C(\underline{b}(\theta)) + C(\underline{g}^*)][C(\underline{b}^*) + C(\underline{g}^*)]^{-1} [C(\underline{b}(\theta)) + C(\underline{g}^*)].$$

By Lehmann & Scheffé (1950, p.323) this is equivalent with $C(\underline{b}^*) + C(\underline{g}^*)$ less than $C(\underline{b}^{**}) + C(\underline{g}^*)$, which is finally equivalent with $C(\underline{b}^*)$ less than $C(\underline{b}^{**})$.

Q E D.

Theorem 4.2. Let \underline{g}^* and $\underline{g}^{**} \in B_2$ and $\underline{b}^* \in B_1$ with

$EC(\underline{b}^*|\theta) \neq 0$. Then $\Pi(\underline{b}^*, \underline{g}^*)$ is less than $\Pi(\underline{b}^*, \underline{g}^{**})$ if and only if $C(\underline{g}^*)$ is less than $C(\underline{g}^{**})$.

Proof. From (4.7c) follows that $\Pi(\underline{b}^*, \underline{g}^*)$ is less than $\Pi(\underline{b}^*, \underline{g}^{**})$ if and only if

$$EC(\underline{b}^*|\theta) [C(\underline{b}^*) + C(\underline{g}^{**})]^{-1} EC(\underline{b}^*|\theta) \text{ is less than}$$

$EC(\underline{b}^*|\theta) [C(\underline{b}^*) + C(\underline{g}^*)]^{-1} EC(\underline{b}^*|\theta)$. By Lehmann & Scheffé (1950, p. 323) this is equivalent with $C(\underline{b}^*) + C(\underline{g}^*)$ less than $C(\underline{b}^*) + C(\underline{g}^{**})$, which is finally equivalent with $C(\underline{g}^*)$ less than $C(\underline{g}^{**})$.

Q E D.

Remarks.

- i) When $q = 1$, the theorems say that we are to prefer the \underline{b}^* and \underline{g}^* with the least variance.

- ii) From (4,6) we see that minimizing $C(\underline{b}^*)$ is equivalent with minimizing $EC(\underline{b}^*|\theta)$.
- iii) When $\underline{\beta}^* = \underline{\beta}$, $\underline{\beta}^*$ is of course independent of every θ -unbiased estimator of $\underline{b}(\theta)$, so that generally in the inhomogeneous case we ought to minimize $C(\underline{b}^*)$ (or equivalently $EC(\underline{b}^*|\theta)$).
- iv) If $EC(\underline{b}^*|\theta) = \underline{0}$, we have $\Pi(\underline{b}^*, \underline{\beta}^*) = \underline{0}$ for all $\underline{\beta}^* \in B_2$. Hence the assumption in Theorem 4.2 that $EC(\underline{b}^*|\theta) \neq \underline{0}$, is necessary.

4E. Before we go further, we shall state some definitions.

Let \underline{x} be a random vector whose distribution depends on some non-random unknown parameter vector $\underline{\beta}$, which is element in a parameter space P , and let $\underline{w} = \underline{w}(\underline{x})$ be a vector function of \underline{x} .

Definition 4.1. We shall say that \underline{w} is sufficient for $\underline{\beta}$ relative to \underline{x} if the conditional distribution of \underline{x} given \underline{w} does not depend on $\underline{\beta}$.

Definition 4.2. We shall say that \underline{x} is complete for $\underline{\beta}$ if $E(f(\underline{x})) = 0$ for all values of $\underline{\beta} \in P$ and a real-valued measurable function f implies that $f(\underline{x}) = 0$ a.s. .

Let θ be an unknown random parameter.

Definition 4.3. We shall say that \underline{w} is θ -sufficient relative to \underline{x} if the conditional distribution of \underline{x} given \underline{w} and θ does not depend on θ .

Example 4.1. Consider an insurance policy that has been running for n insurance terms. In i -th term there have been x_{1i} claims with a total claim amount x_{2i} . $(x_{11}, x_{21}), \dots, (x_{1n}, x_{2n})$ are conditionally independent and identically distributed given an unknown random parameter θ . Given x_{11} and θ

x_{21} has cumulative distribution $G^{x_{11}*}$, where G^{i*} denotes the i -th convolution of a cumulative distribution G , that is, we assume that when a claim has occurred, its distribution is independent of other claims and θ .

Let $\underline{x}_1 = (x_{11}, \dots, x_{1n})'$, $\underline{x}_2 = (x_{21}, \dots, x_{2n})'$, $\underline{x} = (\underline{x}_1, \underline{x}_2)$, and $F(\cdot | \underline{x}_1, \theta)$ the cumulative distribution of \underline{x} given \underline{x}_1 and θ . We easily see that

$$dF(\underline{t} | \underline{x}_1, \theta) = I(\underline{t}_1 = \underline{x}_1) \prod_{i=1}^n dG^{x_{1i}*}(t_{2i}),$$

where I is an indicator. Thus the cumulative distribution of \underline{x} given \underline{x}_1 and θ does not depend on θ , and \underline{x}_1 is θ -sufficient for \underline{x} .

□

Definition 4.4. We shall say that \underline{x} is θ -complete if

$E(f(\underline{x}) | \theta) = 0$ for a real-valued measurable function f implies that $f(\underline{x}) = 0$ a.s. .

4F. In the model of subsection 1J we can write

$$(4.8) \quad \hat{\underline{\beta}} = (\underline{Y}' \underline{\Sigma}^{-1} \underline{Y})^{-1} (\underline{Y}'_1 \underline{\Sigma}_1^{-1} \underline{Y}_1 \hat{\underline{b}} + \underline{Y}'_2 \underline{\Sigma}_2^{-1} \underline{Y}_2 \hat{\underline{\beta}}_2)$$

with

$$\hat{\underline{\beta}}_2 = (\underline{Y}'_2 \underline{\Sigma}_2^{-1} \underline{Y}_2)^{-1} \underline{Y}'_2 \underline{\Sigma}_2^{-1} \underline{x}_2$$

if \underline{Y}_2 has rank q .

From (1.35) and (4.8) we get

$$\begin{aligned} \ddot{\underline{m}} &= \underline{a}' [\underline{Z} \hat{\underline{b}} + (\underline{I} - \underline{Z}) \hat{\underline{\beta}}] = \\ &= \underline{a}' \{ [\underline{Z} + (\underline{I} - \underline{Z}) (\underline{Y}' \underline{\Sigma}^{-1} \underline{Y})^{-1} \underline{Y}'_1 \underline{\Sigma}_1^{-1} \underline{Y}_1] \hat{\underline{b}} + \\ &+ (\underline{I} - \underline{Z}) (\underline{Y}' \underline{\Sigma}^{-1} \underline{Y})^{-1} \underline{Y}'_2 \underline{\Sigma}_2^{-1} \underline{Y}_2 \hat{\underline{\beta}}_2 \} . \end{aligned}$$

By putting

$${}^0 \underline{Z} = \underline{Z} + (\underline{I} - \underline{Z}) (\underline{Y}' \underline{\Sigma}^{-1} \underline{Y})^{-1} \underline{Y}'_1 \underline{\Sigma}_1^{-1} \underline{Y}_1$$

and using the unbiasedness of $\hat{\underline{b}}$ we get

$$\hat{\underline{m}} = \underline{a}' [{}^0\underline{Z} \hat{\underline{b}} + (\underline{I} - {}^0\underline{Z}) \hat{\underline{\beta}}_2].$$

Here $\hat{\underline{b}}$ is a θ -unbiased estimator of $\underline{b}(\theta)$; $\hat{\underline{\beta}}_2$ is an unbiased estimator of $\underline{\beta}$, and $\hat{\underline{\beta}}_2$ is independent of $\hat{\underline{b}}$ and θ .

4G. Now to be more general, let \underline{x}_1 and \underline{x}_2 be two observable random vectors and θ an unknown random parameter. It is assumed that \underline{x}_2 is independent of \underline{x}_1 and θ . We want to estimate $\underline{m} = \underline{a}'\underline{b}(\theta)$, where \underline{a} is a non-random 1×1 -vector and $\underline{b}(\theta)$ a $q \times 1$ vector function of θ with expectation $\underline{\beta}$. We assume that $C(\underline{b}(\theta))$ is invertible. We want our estimator to be of the form $\hat{\underline{m}}(\underline{b}^*, \underline{\beta}^*)$, where \underline{b}^* is a θ -unbiased estimator of $\underline{b}(\theta)$ based on \underline{x}_1 and $\underline{\beta}^*$ an unbiased estimator of $\underline{\beta}$ based on \underline{x}_2 . It is assumed that such estimators exist. We are now in the situation of subsection 4D and can use the theorems there to optimize with respect to \underline{b}^* and $\underline{\beta}^*$.

Now suppose that $\underline{\beta}^*$ is an unbiased estimator of $\underline{\beta}$ based on \underline{x}_2 . If \underline{w}_2 is sufficient for $\underline{\beta}$ relative to \underline{x}_2 , $\check{\underline{\beta}} = E(\underline{\beta}^* | \underline{w}_2)$ is also an unbiased estimator of $\underline{\beta}$, and we would prefer $\check{\underline{\beta}}$ to $\underline{\beta}^*$ in our credibility estimator because $C(\underline{\beta}^*) - C(\check{\underline{\beta}}) = E(C(\underline{\beta}^* | \underline{w}_2))$ is non-negative definite. Hence we can restrict our search for an estimator of $\underline{\beta}$ based on \underline{x}_2 to estimators based on \underline{w}_2 . Furthermore, if \underline{w}_2 is complete for $\underline{\beta}$, $\check{\underline{\beta}}$ is a best unbiased estimator of $\underline{\beta}$ to use in our credibility estimator, because if $\check{\check{\underline{\beta}}}$ is another unbiased estimator of $\underline{\beta}$ based on \underline{w}_2 , we have $E(\check{\underline{\beta}} - \check{\check{\underline{\beta}}}) = 0$ for all values of $\underline{\beta}$ and hence $\check{\underline{\beta}} = \check{\check{\underline{\beta}}}$ a.s.

Now suppose that \underline{b}^* is a θ -unbiased estimator of $\underline{b}(\theta)$ based on \underline{x}_1 . If \underline{w}_1 is θ -sufficient relative to \underline{x}_1 , $\check{\underline{b}} = E(\underline{b}^* | \underline{w}_1)$ is also a θ -unbiased estimator of $\underline{b}(\theta)$ because

$E(\check{\underline{b}}|\theta) = E(E(\underline{b}^*|\underline{w}_1)|\theta) = E(E(\underline{b}^*|\underline{w}_1, \theta)|\theta) = E(\underline{b}^*|\theta) = \underline{b}(\theta)$,
 and we would prefer $\check{\underline{b}}$ to \underline{b}^* in our credibility estimator,
 Because $C(\underline{b}^*) - C(\check{\underline{b}}) = EC(\underline{b}^*|\underline{w}_1)$ is non-negative definite,
 Hence we can restrict our search for a θ -unbiased estimator of $\underline{b}(\theta)$ based on \underline{x}_1 to estimators based on \underline{w}_1 . Furthermore, if \underline{w}_2 is θ -complete, $\check{\underline{b}}$ is a best θ -unbiased estimator of $\underline{b}(\theta)$ to use in our credibility estimator, because if $\check{\underline{b}}$ is another θ -unbiased estimator of $\underline{b}(\theta)$ based on \underline{w}_1 , we have $E(\check{\underline{b}} - \check{\underline{b}}|\theta) = 0$, and hence $\check{\underline{b}} = \check{\underline{b}}$ a.s.

It is interesting to note how independent the search for an optimal \underline{b}^* is of the unconditional distribution of \underline{x}_1 if we know the conditional distribution of \underline{x}_1 given θ .

Example 4.2. Suppose that $\underline{x} = (x_1, \dots, x_r)'$, where x_1, \dots, x_r are claim numbers of an insurance policy with an unknown random risk parameter θ . We assume that x_1, \dots, x_r are conditionally independent and identically Poisson-distributed with expectation θ given θ . We know $\beta = E(\theta) = EC(x_1|\theta)$ and $C(\theta)$ (or at least we are able to estimate them from our portfolio data), but we do not have any further knowledge about the distribution of θ .

We want to estimate θ with an estimator of the form $\check{\theta}(\theta^*, \beta)$ where θ^* is a θ -unbiased estimator of θ . We have that

$\bar{x} = \frac{1}{r} \sum_{i=1}^r x_i$ is θ -sufficient relative to \underline{x} . Furthermore \bar{x} is θ -complete. (This follows e.g. from a general result on completeness in regular Darms-Koopman families.) As \bar{x} is a θ -unbiased estimator of θ , it follows that our optimal estimator is $\check{\theta}(\bar{x}, \beta)$. (If we make the further assumption that θ is Γ -distributed, we have in fact that $\check{\theta}(\bar{x}, \beta) = E(\theta|\underline{x})$.)

We have till now assumed that we ought to use the $\underline{\underline{b}}^*$ and $\underline{\underline{g}}^*$ with the least covariance matrices. However, this may not be so in practice. The estimators with least covariance matrices may be functions of other unknown parameters. And even if we have found $\underline{\underline{b}}^*$ and $\underline{\underline{g}}^*$ not using any unknown parameters, $Z(\underline{\underline{b}}^*, \underline{\underline{g}}^*)$ may be unknown. In practice we usually have to estimate $Z(\underline{\underline{b}}^*, \underline{\underline{g}}^*)$. But then in our choice of $\underline{\underline{b}}^*$ and $\underline{\underline{g}}^*$ we have to take care that it will not be too complicated to estimate $Z(\underline{\underline{b}}^*, \underline{\underline{g}}^*)$. Furthermore our knowledge of the distribution of $\underline{\underline{x}}$ may be so vague that we are not able to find the best $\underline{\underline{b}}^*$ and $\underline{\underline{g}}^*$.

We shall look at two examples.

Example 4.3. Suppose that the observable random vectors $\underline{\underline{x}}_1, \dots, \underline{\underline{x}}_n$ are conditionally independent and identically distributed given an unknown random variable θ . We want to estimate $m = \underline{\underline{a}}' \underline{\underline{b}}(\theta)$, where $\underline{\underline{a}}$ is a non-random $q \times 1$ -vector and $\underline{\underline{b}}(\theta)$ a $q \times 1$ vector function of θ with expectation $\underline{\underline{g}}$. We assume that $C(\underline{\underline{b}}(\theta))$ is invertible. Our estimator is to be of the form $m(\underline{\underline{b}}^*, \underline{\underline{g}})$ where $\underline{\underline{b}}^*$ is some θ -unbiased estimator of $\underline{\underline{b}}(\theta)$.

If there exists a θ -unbiased estimator $\check{\underline{\underline{b}}}$ of $\underline{\underline{b}}(\theta)$ based on a θ -sufficient (relative to $\underline{\underline{x}}_1, \dots, \underline{\underline{x}}_n$) and θ -complete statistic, this estimator is of course optimal in least mean square sense. However, if $Z(\check{\underline{\underline{b}}}, \underline{\underline{g}})$ is to be estimated by the present data and similar data from other realizations of θ , this may be rather messy unless $\check{\underline{\underline{b}}}$ is on the form $\check{\underline{\underline{b}}} = \frac{1}{n} \sum_{i=1}^n \underline{\underline{b}}^*(\underline{\underline{x}}_i)$. Here $\underline{\underline{b}}^*(\underline{\underline{x}}_1)$ is of course a θ -unbiased estimator of $\underline{\underline{b}}(\theta)$.

We shall now restrict $\underline{\underline{b}}^*$ to be of the form

$$\underline{\underline{b}}^* = \frac{1}{n} \sum_{i=1}^n \underline{\underline{b}}^*(\underline{\underline{x}}_i).$$

Then we have

$$(4.9) \quad C(\underline{b}^*) = C(\underline{b}(\theta)) + \frac{1}{n} EC(\underline{b}^*(\underline{x}_1) | \theta),$$

Suppose now that there exists a θ -unbiased estimator $\underline{b}^*(\underline{x}_1)$ based on a θ -sufficient (relative to \underline{x}_1) and θ -complete statistic. From (4.9) follows then that

$$\underline{\bar{b}} = \frac{1}{n} \sum_{i=1}^n \underline{b}(\underline{x}_i)$$

is our optimal estimator. We get

$$\begin{aligned} Z(\underline{\bar{b}}, \underline{g}) &= C(\underline{b}(\theta)) [C(\underline{b}(\theta)) + \frac{1}{n} EC(\underline{b}^*(\underline{x}_1) | \theta)]^{-1} = \\ n [n \underline{I} + EC(\underline{b}^*(\underline{x}_1) | \theta) C(\underline{b}(\theta))^{-1}]^{-1} &= n (n \underline{I} + \underline{K})^{-1} \end{aligned}$$

with

$$\underline{K} = EC(\underline{b}^*(\underline{x}_1) | \theta) C(\underline{b}(\theta))^{-1}$$

giving

$$\hat{m}(\underline{\bar{b}}, \underline{g}) = \underline{a}' [n (n \underline{I} + \underline{K})^{-1} \underline{\bar{b}} + \underline{K} (n \underline{I} + \underline{K})^{-1} \underline{g}].$$

Example 4.4. Let $(\underline{x}_1', m_1, \theta_1)'$, ..., $(\underline{x}_n', m_n, \theta_n)'$ be independent random vectors. $\theta_1, \dots, \theta_n$ are unknown and identically distributed. For each i \underline{x}_i is an observable random vector and m_i an unknown random variable, \underline{x}_i and m_i are independent given θ_i , and

$$E(m_i | \theta_i) = \underline{a}_i' \underline{b}(\theta_i).$$

Here \underline{a}_i is a non-random $q \times 1$ -vector and $\underline{b}(\theta_i)$ a $q \times 1$ vector function of θ_i with unknown expectation \underline{g} . We assume that $C(\underline{b}(\theta_1))$ is invertible. The conditional distribution of \underline{x}_i given θ_i is known, but the distribution of θ_i is unknown.

We want to estimate m_k with an estimator of the form $\hat{m}_k(\underline{b}_k^*, \underline{g}^*)$, where \underline{b}_k^* is a θ_k -unbiased estimator of $\underline{b}(\theta_k)$, and \underline{g}^* is an unbiased estimator of \underline{g} .

Assume that for each i there exists a θ_i -unbiased estimator \check{b}_i of $b(\theta_i)$ such that for any other θ_i -unbiased estimator b_i^* of $b(\theta_i)$ we have $C(b_i^*) - C(\check{b}_i)$ non-negative definite.

By comparing with (1.43)

$$a_k' \{ Z(\check{b}_k, \underline{g}) \check{b}_k + [I - Z(\check{b}_k, \underline{g})] [(\sum_{i=1}^n Z(\check{b}_i, \underline{g}))^{-1} \sum_{i=1}^n Z(\check{b}_i, \underline{g}) \check{b}_i] \}$$

seems to be a natural estimator of m . From subsection 4F follows that this estimator can be written on the form $\check{m}_k(\check{b}_k, \underline{g}^*)$ with

$$\underline{g}_k^* = [\sum_{i \neq k} Z(\check{b}_i, \underline{g})]^{-1} \sum_{i \neq k} Z(\check{b}_i, \underline{g}) \check{b}_i,$$

and is thereby of the required form.

□

4H. The inhomogeneous case ($\underline{g}^* = \underline{g}$) has previously been treated by Pechlivanides (1973) and Taylor (1977). Taylor's treatment is very similar to the present one.

5. Optimal semilinear estimators.

Conditions are given under which exchangeability implies conditional independence. De Vylder's (1976b) theory of optimal semilinear estimators is slightly generalized. A condition is given under which optimal semilinear estimators can be developed recursively. Some asymptotic properties of optimal semilinear estimators are discussed. Conditions are given under which optimal semilinear estimators can be based on θ -sufficient estimators.

5A. Let $\{\underline{x}_i\}_{i=1}^{\infty}$ be a sequence of real random $r \times 1$ -vectors and m a real random variable. We assume that $\underline{x}_1, \dots, \underline{x}_n$ are exchangeable relative to m for all n , and that $E(m^2) < \infty$. Let S be the set of possible outcomes of $(m, \underline{x}_1, \underline{x}_2, \dots)$, A the σ -field of Borel sets in S , and P the probability measure induced on (S, A) by $(m, \underline{x}_1, \underline{x}_2, \dots)$.

From a trivial multi-dimensional generalization of a result in Loève (1963, p.365) follows that $\underline{x}_1, \underline{x}_2, \dots$ are conditionally independent and identically distributed given a sub- σ -field A' of A . Furthermore by Loève (1963, p.363) follows that the conditional probability measure $P^{A'}$ on (S, A) given A' can be regularized. Hence we may assume that $P^{A'}$ is regular.

Lemma 5.1. Let $f : \mathbb{R}^{rs} \rightarrow \mathbb{R}$ and $g : \mathbb{R} \rightarrow \mathbb{R}$ be bounded Baire functions. Then

$$C(g(m), f(\underline{x}_1, \dots, \underline{x}_s) | A') = 0.$$

Proof. For all n we have

$$\begin{aligned} 0 &\leq C(n^{\frac{1}{2}} g(m) + \sum_{j=0}^{n-1} f(\underline{x}_{js+1}, \dots, \underline{x}_{(j+1)s}) | A') = \\ &n C(g(m) | A') + n C(f(\underline{x}_1, \dots, \underline{x}_s) | A') + 2 n^{\frac{3}{2}} C(g(m), f(\underline{x}_1, \dots, \underline{x}_s) | A'), \end{aligned}$$

giving

$$0 \leq n^{-\frac{1}{2}} (C(g(m) | A') + C(f(\underline{x}_1, \dots, \underline{x}_s) | A')) + 2 C(g(m), f(\underline{x}_1, \dots, \underline{x}_s) | A')$$

By letting $n \rightarrow \infty$ we get

$$(5.1) \quad C(g(m), f(\underline{x}_1, \dots, \underline{x}_s) | A') \geq 0.$$

A similar treatment of

$$C(n^{\frac{1}{2}} g(m) - \sum_{j=0}^{n-1} f(\underline{x}_{js+1}, \dots, \underline{x}_{(j+1)s}) | A') \text{ gives}$$

$$(5.2) \quad C(g(m), f(\underline{x}_1, \dots, \underline{x}_s) | A') \leq 0.$$

Lemma 5.1 now follows from (5.1) and (5.2).

Q E D.

Theorem 5.1. m is conditionally independent of $\{\underline{x}_i\}_{i=1}^{\infty}$ given A' .

Proof. Let A be a Borel set in \mathbb{R} and A_1, \dots, A_n Borel sets in \mathbb{R}^r . It is sufficient to show that

$$P^{A'}((m \in A) \cap \bigcap_{i=1}^n (\underline{x}_i \in A_i)) = P^{A'}(m \in A) P^{A'}(\bigcap_{i=1}^n (\underline{x}_i \in A_i)).$$

By Lemma 5.1

$$0 = C(I(m \in A), I(\bigcap_{i=1}^n (\underline{x}_i \in A_i) | A')) =$$

$$E(I(m \in A) I(\bigcap_{i=1}^n (\underline{x}_i \in A_i) | A')) - E(I(m \in A) | A') E(I(\bigcap_{i=1}^n (\underline{x}_i \in A_i) | A'))$$

$$= P^{A'}((m \in A) \cap \bigcap_{i=1}^n (\underline{x}_i \in A_i)) - P^{A'}(m \in A) P^{A'}(\bigcap_{i=1}^n (\underline{x}_i \in A_i)).$$

Q E D .

5B. Let F be the set of all Baire functions $f : \mathbb{R}^r \rightarrow \mathbb{R}$ such that $E(f(\underline{x}_1)) < \infty$.

For all n let

$$m_n^* = \sum_{i=1}^n f_n^*(\underline{x}_i)$$

be the optimal semilinear estimator of m based on $\underline{x}_1, \dots, \underline{x}_n$,

that is, the best estimator of m of the form $\sum_{i=1}^n f(\underline{x}_i)$ with $f \in F$.

The following theorem, which is a slight extension of a result by De Vylder (1976b, section 7), states that m_n^* always exists and is almost surely unique.

Theorem 5.2. The functional equation

$$(5.3) \quad f(\underline{x}_1) + (n-1) E(f(\underline{x}_2) | \underline{x}_1) = E(m | \underline{x}_1) \text{ a.s.}$$

has always a solution. If both $f^{(1)}$ and $f^{(2)}$ are solutions of (5.3), then $f^{(1)}(\underline{x}_1) = f^{(2)}(\underline{x}_1)$ a.s. . $f_n^* \in F$ satisfies

$$E\left(\sum_{i=1}^n f_n^*(\underline{x}_i) - m\right)^2 \leq E\left(\sum_{i=1}^n f(\underline{x}_i) - m\right)^2$$

for all $f \in F$ if and only if f_n^* satisfies (5.3),

The proof goes almost exactly as the deductions by De Vylder (1976b, sections 4 and 7) (replace θ with A' and X_{t+1} with m), and is therefore omitted.

Remarks.

i) We see that if f_n^* satisfies (5.3), then

$$E(f_n^*(\underline{x}_1)) = \frac{1}{n} E(m). \text{ Hence } m_n^* \text{ is an unbiased estimator of } m.$$

ii) From (5.3) follows that $m_1^* = f_1^*(\underline{x}_1) = E(m | \underline{x}_1)$ a.s. . This result is obvious.

iii) Theorem 5.1 gives

$$E(m | \underline{x}_1) = E(E(m | \underline{x}_1, A') | \underline{x}_1) = E(E(m | A') | \underline{x}_1).$$

From this and (5.3) follows that m_n^* is an optimal semilinear estimator of m if and only if m_n^* is an optimal semilinear estimator of $E(m|A')$.

5C. If we get the \underline{x}_i 's sequentially, it might be interesting to estimate m whenever we get a new \underline{x}_i . (\underline{x}_i could e.g. be the claim data of i -th insurance term of an insurance policy.) Then it would be convenient if f_1^*, f_2^*, \dots could be chosen such that there was an easy connection between m_n^* and m_{n+1}^* for all n (e.g. something like (1.32)). It would be most inconvenient to have to keep all the previous \underline{x}_i 's for use in future estimators.

After these remarks we wish that there exists a version $\{f_n^*\}_{n=1}^{\infty}$ of the optimal functions and a sequence $\{g_n\}_{n=1}^{\infty}$ of functions such that

$$(5.4) \quad m_{n+1}^* = g_n(m_n^*, \underline{x}_{n+1})$$

for all n .

The following theorem gives a condition under which our wish is satisfied.

Theorem 5.3. Assume that there exists a non-negative number π such that

$$(5.5) \quad E(E(m|\underline{x}_2)|\underline{x}_1) = \pi E(m|\underline{x}_1) + (1 - \pi) E(m) \text{ a.s. .}$$

Then an optimal sequence $\{f_n^*\}_{n=1}^{\infty}$ can be defined by

$$(5.6) \quad f_n^*(\underline{x}_1) = \frac{1}{(n-1)\pi + 1} E(m|\underline{x}_1) + \frac{(n-1)(\pi-1)}{(n-1)\pi + 1} \frac{1}{n} E(m).$$

We then have

$$(5.7) \quad m_n^* = \frac{n}{(n-1)\pi + 1} \frac{1}{n} \sum_{i=1}^n E(m|\underline{x}_i) + \frac{(n-1)(\pi-1)}{(n-1)\pi + 1} E(m)$$

and the recursion

$$(5.8) \quad m_{n+1}^* = \frac{(n-1)\pi+1}{n\pi+1} m_n^* + \frac{1}{n\pi+1} E(m|\underline{x}_{n+1}) + \frac{\pi-1}{n\pi+1} E(m).$$

Proof. We have to show that f_n^* defined by (5.6) satisfies (5.3).

We have

$$(5.9) \quad E(f_n^*(\underline{x}_2)|\underline{x}_1) = \frac{1}{(n-1)\pi+1} E(E(m|\underline{x}_2)|\underline{x}_1) + \frac{(n-1)(\pi-1)}{(n-1)\pi+1} \frac{1}{n} E(m).$$

Inserting (5.5) into (5.9) and rearranging give

$$(5.10) \quad E(f_n^*(\underline{x}_2)|\underline{x}_1) = \frac{\pi}{(n-1)\pi+1} E(m|\underline{x}_1) - \frac{\pi-1}{(n-1)\pi+1} \frac{1}{n} E(m) \text{ a.s.}$$

From (5.6) and (5.10) we get

$$f_n^*(\underline{x}_1) + (n-1) E(f_n^*(\underline{x}_2)|\underline{x}_1) = E(m|\underline{x}_1) \text{ a.s.}$$

Hence f_n^* satisfies (5.3), and Theorem 5.2 gives that f_n^* is optimal.

(5.7) follows easily by adding $f_n^*(\underline{x}_1), \dots, f_n^*(\underline{x}_n)$, and (5.8) follows trivially from (5.7).

This completes the proof of Theorem 5.3.

Q E D.

Remark. It can easily be shown that if there exist a version

$\{f_n^*\}_{n=1}^\infty$ of the optimal functions and sequences $\{\mu_n\}_{n=1}^\infty$ and $\{v_n\}_{n=1}^\infty$ of constants such that

$$(5.11) \quad f_{n+1}^*(\underline{x}_1) = \mu_n f_n^*(\underline{x}_1) + v_n$$

for all n , then (5.5) must be satisfied. It seems intuitively plausible that we must have (5.11) whenever (5.4) is satisfied. However, it seems that we have to put restrictions on the distributions of the \underline{x}_i 's and m to be able to prove such a result.

The assumption in Theorem 5.3 that π was non-negative, was made to ensure that the denominator in (5.6) was non-zero. However, the following lemma shows that it is sufficient to assume the existence of a real constant π satisfying (5.5) to be able to define f_{n+1}^* as a linear function of f_n^* .

Lemma 5.3. If there exists a real constant π satisfying (5.5), this π can be chosen non-negative.

Proof. We have

$$\begin{aligned} 0 &\leq CE(E(m|\underline{x}_1)|A') = C(E(m|\underline{x}_1), E(m|\underline{x}_2)) = \\ &C(E(m|\underline{x}_1), E(E(m|\underline{x}_2)|\underline{x}_1)) = \\ &C(E(m|\underline{x}_1), \pi E(m|\underline{x}_1) + (1 - \pi) E(m)) = \pi CE(m|\underline{x}_1). \end{aligned}$$

Hence

$$\pi CE(m|\underline{x}_1) \geq 0.$$

If $CE(m|\underline{x}_1) > 0$, π must be greater than or equal to zero. If $CE(m|\underline{x}_1) = 0$, $E(m|\underline{x}_1) = E(m)$ a.s., and we can let π be any number.

Hence Lemma 5.3 is proved.

Q E D.

5D. Let $m_n^* = \sum_{i=1}^n f_n^*(\underline{x}_i)$ be an optimal semilinear estimator of m . We shall look at some asymptotic results when $n \rightarrow \infty$.

Theorem 5.4. Assume that there exists a function $\dot{f} \in F$ such that $E(\dot{f}(\underline{x}_1)|A') = E(m|A')$. Then

$$m_n^* \xrightarrow{P} E(m|A').$$

If, in addition, there exists a positive constant π such that (5.5) holds, then

$$m_n^* \xrightarrow{a.s.} E(m|A').$$

Proof. As m_n^* is an optimal semilinear estimator of $E(m|A')$, m_n^* is in particular not worse than $\dot{m}_n = \sum_{i=1}^n \frac{1}{n} \dot{f}(x_{\underline{i}})$. This gives

$$0 \leq C(m_n^* - E(m|A')) = E(m_n^* - E(m|A'))^2 \leq E(\dot{m}_n - E(m|A'))^2 =$$

$$EC(\dot{m}_n | A') = \frac{1}{n} EC(\dot{f}(x_{\underline{1}}) | A') \leq \frac{1}{n} C(\dot{f}(x_{\underline{1}})).$$

As $\dot{f} \in F$, $C(\dot{f}(x_{\underline{1}})) < \infty$. From this follows that

$$\frac{1}{n} C(\dot{f}(x_{\underline{1}})) \rightarrow 0 \text{ as } n \rightarrow \infty, \text{ giving}$$

$$C(m_n^* - E(m|A')) \rightarrow 0$$

as $n \rightarrow \infty$. By Chebyshev's inequality this proves the first part of the theorem.

If (5.5) holds, $m_n^* = m_n^{**}$ a.s., where

$$m_n^{**} = \frac{n}{(n-1)\pi+1} \frac{1}{n} \sum_{i=1}^n E(m|x_{\underline{i}}) + \frac{(n-1)(\pi-1)}{(n-1)\pi+1} E(m).$$

The strong law of large numbers gives

$$(5.12) \quad \frac{1}{n} \sum_{i=1}^n E(m|x_{\underline{i}}) \xrightarrow{\text{a.s.}} E(E(m|x_{\underline{1}}) | A').$$

Furthermore

$$(5.13) \quad \frac{n}{(n-1)\pi+1} \rightarrow \frac{1}{\pi}.$$

(5.12) and (5.13) now give

$$m_n^{**} \xrightarrow{\text{a.s.}} \frac{1}{\pi} E(E(m|x_{\underline{1}}) | A') + (1 - \frac{1}{\pi}) E(m),$$

giving

$$m_n^* \xrightarrow{\text{a.s.}} \frac{1}{\pi} E(E(m|x_{\underline{1}}) | A') + (1 - \frac{1}{\pi}) E(m).$$

As almost sure convergence implies convergence in probability,

$$m_n^* \xrightarrow{P} \frac{1}{\pi} E(E(m|x_{\underline{1}}) | A') + (1 - \frac{1}{\pi}) E(m).$$

But by the first part of the theorem

$$m_n^* \xrightarrow{P} E(m|A').$$

Hence

$$\frac{1}{\pi} E(E(m|\underline{x}_1)|A') + (1 - \frac{1}{\pi}) E(m) = E(m|A') \quad \text{a.s.},$$

giving

$$m_n^* \xrightarrow{\text{a.s.}} E(m|A').$$

Q E D.

5E. Suppose now that $\underline{x}_1, \underline{x}_2, \dots$ are conditionally independent and identically distributed given a random variable θ on (S, A, P) . We can now let A' be the σ -field generated by θ . Hence by Theorem 5.1 m is independent of $\{\underline{x}_i\}_{i=1}^{\infty}$ given θ .

We get the following version of Theorem 5.4.

Theorem 5.4'. Assume that there exists a function $f \in F$ such that $f(\underline{x}_1)$ is a θ -unbiased estimator of m . Then

$$m_n^* \xrightarrow{P} E(m|\theta).$$

If, in addition, there exists a positive constant π such that (5.5) holds, then

$$m_n^* \xrightarrow{\text{a.s.}} E(m|\theta).$$

Example 5.1. Assume that the random variables $\underline{x}_1, \underline{x}_2, \dots$ are conditionally independent and identically distributed given an unknown random variable θ , and that $C(\underline{x}_1)$ exists. We want to estimate \underline{x}_{n+1} , with an optimal semilinear estimator based on $\underline{x}_1, \dots, \underline{x}_n$. But this is the same as wanting an optimal semilinear estimator of $m = E(\underline{x}_{n+1}|\theta) = E(\underline{x}_1|\theta)$ based on $\underline{x}_1, \dots, \underline{x}_n$.

This model has been studied by De Vylder and Ballegeer (De Vylder & Ballegeer (1975), De Vylder (1976b)).

As $E(x_1|\theta) = E(m|\theta)$, Theorem 5.4 gives that

$$m_n^* \xrightarrow{P} E(x_1|\theta).$$

Furthermore, if there exists a positive constant π such that $E(E(x_1|\theta)|x_2)|x_1) = \pi E(E(x_1|\theta)|x_1) + (1 - \pi) E(x_1)$ a.s., then

$$m_n^* \xrightarrow{\text{a.s.}} E(x_1|\theta).$$

□

From Theorem 5.4' follows that if we make the two assumptions

i) there exists a function $f \in F$ such that

$$E(f(\underline{x}_1)|\theta) = E(m|\theta);$$

ii) there exists a positive constant π such that (5.5) holds,

$$\text{then } m_n^* \xrightarrow{\text{a.s.}} E(m|\theta).$$

From the proof of Theorem 5.4 follows that whenever assumption ii) is satisfied,

$$(5.14) \quad m_n^* \xrightarrow{\text{a.s.}} \frac{1}{\pi} E(E(m|\underline{x}_1)|\theta) + \left(1 - \frac{1}{\pi}\right) E(m),$$

and if, in addition, assumption i) is satisfied,

$$(5.15) \quad \frac{1}{\pi} E(E(m|\underline{x}_1)|\theta) + \left(1 - \frac{1}{\pi}\right) E(m) = E(m|\theta) \quad \text{a.s.}$$

If both (5.14) and (5.15) are satisfied, $m_n^* \xrightarrow{\text{a.s.}} E(m|\theta)$. A natural question is then: Is it necessary to make assumption i), or does assumption ii) generally imply (5.15), such that $m_n^* \xrightarrow{\text{a.s.}} E(m|\theta)$ whenever assumption ii) is satisfied?

The following example gives a case where assumption ii) is satisfied, and m_n^* does not converge towards $E(m|\theta)$. Hence our question is answered.

Example 5.2. The random variables x_1, x_2, \dots are conditionally independent and identically distributed given an unknown random variable θ , and the random variable m is independent of the x_i 's given θ . We assume that there exists a positive constant π satisfying

$$E(E(m|x_2)|x_1) = \pi E(m|x_1) + (1 - \pi) E(m) \quad \text{a.s.},$$

and that

$$E(x_1|\theta) = E(m|\theta) = b(\theta).$$

From Theorem 5.4' follows that $m_n^* \xrightarrow{\text{a.s.}} b(\theta)$.

Let us now make the further assumption that

$$m = b(\theta) + \eta,$$

where the random variable η has expectation zero and positive variance, and is independent of the x_i 's and θ . Let ${}^0\theta = (\theta, \eta)$. We now have that x_1, x_2, \dots are conditionally independent and identically distributed given ${}^0\theta$, and m is independent of the x_i 's given ${}^0\theta$. Assumption ii) is satisfied, but as

$m_n^* \xrightarrow{\text{a.s.}} b(\theta)$, m_n^* cannot converge almost surely towards $E(m|{}^0\theta) = b(\theta) + \eta$.

□

Theorem 5.5. Let \underline{w} be a Baire function from \mathbb{R}^r to \mathbb{R}^p , and assume that $\underline{w}(\underline{x}_1)$ is θ -sufficient for \underline{x}_1 . Then for all n there exists a Baire function h_n from \mathbb{R}^p to \mathbb{R} such that

$\sum_{i=1}^n h_n(\underline{w}(\underline{x}_i))$ is an optimal semilinear estimator of m based

on $\underline{x}_1, \dots, \underline{x}_n$.

Proof. Let $m_n^* = \sum_{i=1}^n f_n^*(x_i)$ be an optimal semilinear estimator of m and thereby of $E(m|\theta)$. Let $h_n(w(x_1)) = E(f_n^*(x_1)|w(x_1))$

and $m_n^{**} = \sum_{i=1}^n h_n(w(x_i))$. We have

$$(5.16) \quad E(m_n^* - E(m|\theta))^2 = EE(\{[m_n^{**} - E(m|\theta)] + (m_n^* - m_n^{**})\}^2 | \theta, w(x_1), \dots, w(x_n)).$$

We get

$$\begin{aligned} E([m_n^{**} - E(m|\theta)] (m_n^* - m_n^{**}) | \theta, w(x_1), \dots, w(x_n)) &= \\ [m_n^{**} - E(m|\theta)] E(m_n^* - m_n^{**} | \theta, w(x_1), \dots, w(x_n)) &= \\ [m_n^{**} - E(m|\theta)] (m_n^{**} - m_n^{**}) &= 0. \end{aligned}$$

Thus when multiplying out the right-hand side of (5.16), the product term vanishes, and we get

$$E(m_n^* - E(m|\theta))^2 = E(m_n^{**} - E(m|\theta))^2 + E(m_n^* - m_n^{**})^2.$$

Thus

$$E(m_n^* - E(m|\theta))^2 \geq E(m_n^{**} - E(m|\theta))^2.$$

But m_n^* was an optimal semilinear estimator of $E(m|\theta)$ and m_n^{**} a semilinear estimator of $E(m|\theta)$. Then m_n^{**} must be an optimal semilinear estimator of $E(m|\theta)$ and consequently an optimal semilinear estimator of m .

This proves Theorem 5.5.

Q E D.

References.

- Bailey, A.L. (1942). Sampling Theory in Casualty Insurance.
Proceedings of the Casualty Actuarial Society 29, pp. 50-93.
- Bailey, A.L. (1945). A Generalized Theory of Credibility.
Proceedings of the Casualty Actuarial Society 32, pp. 13-20.
- Bühlmann, H. (1967). Experience Rating and Credibility.
ASTIN Bulletin 4, pp. 199-207.
- Bühlmann, H. & Straub, E. (1970). Glaubwürdigkeit für
Schadensätze. Mitteilungen der Vereinigung Schweizerischer
Versicherungsmathematiker 70, pp. 111-133. Also appears as:
Credibility for Loss Ratios. ARCH 1972.2. Translation by
C.E. Brooks.
- De Vylder, F. (1976a). Geometrical Credibility. Scandinavian
Actuarial Journal 1976, pp. 121-149.
- De Vylder, F. (1976b). Optimal Semilinear Credibility.
Mitteilungen der Vereinigung Schweizerischer Versicherungs-
mathematiker 76, pp. 27-39.
- De Vylder, F. (1978a). Parameter Estimation in Credibility
Theory. ASTIN Bulletin 10, pp. 99-112.
- De Vylder, F. (1978b). Optimal Parameter Estimation in Semi-
distributionfree Credibility Theory. To be published in
ASTIN Bulletin.
- De Vylder, F. & Ballegeer, Y. (1975). A Numerical Illustration
of Optimal Semilinear Credibility. University of Louvain,
Belgium. To be published in ASTIN Bulletin.

- Gerber, H.U. & Jones, D.A. (1973). Credibility Formulae with Geometric Weights. Proceedings of the Business and Economics Section. American Statistical Association 1973.
- Gerber, H.U. & Jones, D.A. (1975). Credibility Formulas of the Updating Type. Transactions of the Society of Actuaries 27, pp. 31-46. Also appears in Kahn (1975), pp. 89-105.
- Hachemeister, C.A. (1975). Credibility for Regression Models with Application to Trend. Kahn (1975), pp. 129-163.
- Hewitt, C.C., Jr. (1970). Credibility for Severity. Proceedings of the Casualty Actuarial Society 57, pp. 148-171. Discussion. Proceedings of the Casualty Actuarial Society 58, pp. 25-31.
- Jewell, W.S. (1975a). Bayesian Regression and Credibility Theory. RM-75-63, International Institute for Applied Systems Analysis, Laxenburg, Austria. Also appears in Jewell (1976b).
- Jewell, W.S. (1975b). Credibility Regression with Simple Trends. ORC 75-16, Operations Research Center, University of California, Berkeley.
- Jewell, W.S. (1975c). Model Variations in Credibility Theory. Kahn (1975), pp. 193-244.
- Jewell, W.S. (1975d). The Use of Collateral Data in Credibility Theory: A Hierarchical Model. RM-75-24, International Institute for Applied Systems Analysis, Laxenburg, Austria. Also appears in Jewell (1976b).
- Jewell, W.S. (1976a). A Survey of Credibility Theory. ORC 76-31, Operations Research Center, University of California, Berkeley.

Jewell, W.S. (1976b). Three Papers in Credibility Theory.

ORC 76-16, Operations Research Center, University of California, Berkeley.

Jewell, W.S. & Avenhaus, R. (1975). Bayesian Inverse

Regression and Discrimination: An Application of Credibility Theory. RM-75-27. Institute of Applied Systems Analysis, Laxenburg, Austria. Also appears in Jewell (1976b).

Kahn, P.M. (ed.) (1975). Credibility: Theory and Applications.

Proceedings of Actuarial Research Conference on Credibility, Berkeley, September 1974. Academic Press, New York.

Lehmann, E.L. & Scheffé, H. (1950). Completeness, Similar

Regions, and Unbiased Estimation. Part 1. Sankhyā 10, pp. 305-340.

Loève, M. (1963). Probability Theory. 3rd ed. Van Nostrand, Princeton, New Jersey.

Longley-Cook, L.H. (1962). An Introduction to Credibility

Theory. Proceedings of the Casualty Actuarial Society 49, pp. 194-221.

Norberg, R. (1978). Empirical Bayes Credibility. Submitted for publication in Scandinavian Actuarial Journal.

Pechlivanides, P.M. (1973). Approximations by Orthogonal

Functions in Casualty Insurance. ORC 73-24. Operations Research Center, University of California, Berkeley.

Taylor, G.C. (1974). Experience Rating with Credibility Adjustment of the Manual Premium. ASTIN Bulletin 7, pp. 323-336.

Taylor, G.C. (1975). Credibility for Time-Heterogeneous Loss Ratios. Kahn (1975), pp. 363-390.

Taylor, G.C. (1977). Abstract Credibility. Scandinavian Actuarial Journal 1977, pp. 149-168.

Welten, C.P. (1968). The Unearned No Claim Bonus. ASTIN Bulletin 5, pp. 25-32.

Whitney, A.W. (1918). The Theory of Experience Rating. Proceedings of the Casualty Actuarial Society 4, pp. 274-292.

Zellner, A. (1971). An Introduction to Bayesian Inference in Econometrics. Wiley, New York.